COMPUTATIONAL AND EXPERIMENTAL INVESTIGATIONS OF FORCES IN PROTEIN FOLDING

A Dissertation

by

DAVID ANDREW SCHELL

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2003

Major Subject: Biochemistry

COMPUTATIONAL AND EXPERIMENTAL INVESTIGATIONS OF FORCES IN PROTEIN FOLDING

A Dissertation

by

DAVID ANDREW SCHELL

Submitted to Texas A&M University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Approved as to style and content by:

J. Martin Scholtz (Chair of Committee)	Gary R. Kunkel (Member)
C. Nick Pace (Member)	Andy C. LiWang (Member)
	Gregory D. Reinhart (Head of Department)

December 2003

Major Subject: Biochemistry

ABSTRACT

Computational and Experimental Investigations of Forces in Protein Folding.

(December 2003)

David Andrew Schell, B.S. Texas A&M University

Chair of Advisory Committee: Dr. J. Martin Scholtz

Properly folded proteins are necessary for all living organisms. Incorrectly folded proteins can lead to a variety of diseases like Alzheimer's Disease or Bovine Spongiform Encephalitis (Mad Cow Disease). Understanding the forces involved in protein folding is essential to the understanding and treatment of protein misfolding diseases. When proteins fold, a significant amount of surface area is buried in the protein interior. It has long been known that burial of hydrophobic surface area was important to the stability of the folded structure. However, the impact of burying polar surface area is not well understood. Theoretical results suggest that burying polar groups decreases the stability, but experimental evidence supports the belief that polar group burial increases the stability. Studies of tyrosine to phenylalanine mutations have shown the removal of the tyrosine OH group generally decreases stability. Through computational investigations into the effect of buried tyrosine on protein stability, favorable van der Waals interactions are shown to correlate with the change in stability caused by replacing the tyrosine with phenylalanine to remove the polar OH group. Two large-scale studies on nearly 1000 high-resolution x-ray structures are presented. The first investigates the electrostatic and van der Waals

interactions, analyzing the energetics of burying various atom groups in the protein interior. The second large-scale study analyzes the packing differences in the interior of the protein and shows that hydrogen bonding increases packing, decreasing the volume of a hydrogen bonded backbone by about 1.5 $Å^3$ per hydrogen bond. Finally, a structural comparison between RNase Sa and a variant in which five lysines replaced five acidic groups to reverse the net charge is presented. It is shown that these mutations have a marginal impact on the structure, with only small changes in some loop regions.

DEDICATION

To my wife Margaret

whom I love with all my heart

and our daughter Cora

my greatest accomplishment.

ACKNOWLEDGMENTS

I would like to thank my committee: Marty Scholtz, my committee chair, Gary Kunkel, Andy LiWang, and Nick Pace. Marty Scholtz and Nick Pace have provided guidance and friendship throughout my career at Texas A&M University and Andy LiWang provided assistance with the NMR.

I thank Jerry Tsai for his help with the computational studies and for all the helpful comments about my dissertation. I also thank Dr. Karl Koshlap and the Biomolecular NMR Laboratory at Texas A&M University for use of the NMR facilities and assistance in collecting the NMR data.

I would also like to thank all the members of the Pace and Scholtz Labs, past and present, who have provided help, friendship, and stimulating conversation over the many years.

TABLE OF CONTENTS

Page

ABSTRACT iii

DEDICATION v

ACKNOWLEDGMENTS vi

TABLE OF CONTENTS vii

LIST OF TABLES ix

LIST OF FIGURES xi

INTRODUCTION 1

- Thermodynamic Hypothesis 1
- Historical View of Protein Folding 2
- Conformational Stability 3
- Forces in Protein Folding 4
- RNase Sa and Variants 16

TYROSINE HYDROXYLS 21

- Introduction 21
- Methods 25
- Results 31
- Discussion 38
- Conclusions 41

ELECTROSTATIC AND VAN DER WAALS INTERACTIONS 43

- Introduction 43
- Methods 45
- Results and Discussion 48
- Conclusion 75

PACKING OF BURIED ATOMS 81

- Introduction 81
- Methods 83
- Results 89

Page

Discussion 100
Conclusion 102

STRUCTURAL COMPARISON OF RNASE SA AND 5K VARIANT 103

Introduction 103
Methods 106
Results and Discussion 107
Conclusion 122

SUMMARY 124

REFERENCES 127

VITA 140

LIST OF TABLES

TABLE		Page
1	Burial of tyrosine hydroxyl	25
2	Parameters for amino acids	26
3	Calculation of total contributions of tyrosine hydroxyl group from molecule A of 1lni	32
4	Correlation coefficients between calculated terms and observed ΔΔG	35
5	Variation in AMBER calculations between different structures of RNase Sa	35
6	Interaction of tyrosine hydroxyl group to atoms within 3.5 Å of the Oη	37
7	Protein data set (911 Structures)	46
8	Classification of atoms	48
9	Electrostatic interactions between the given atom group and class	55
10	Lennard-Jones attractive interactions between the given atom group and class	63
11	Lennard-Jones repulsive interactions between given atom group and class	71
12	Summary of significant interactions	76
13	Chothia radii	84
14	Protein data set (872 Files)	85
15	Differences in residue volume associated with solvent exposure	90
16	Differences in volume associated with hydrogen bonding in buried residues	92
17	Volume changes associated with differences in buried secondary structures	94

TABLE Page

18 Packing of atom groups associated with solvent exposure 95

19 Differences in packing due to hydrogen bonding and secondary structure 97

20 Comparison of residue volumes to other studies 101

21 Comparison of crystal structures 108

22 Comparison of secondary structure 109

23 RMSD between RNase Sa and the 5K variant 111

24 Backbone amide hydrogen bonds 114

25 Side chain donor hydrogen bonds 115

26 Observed trans hydrogen bond J-coupling 119

27 Residues volumes in RNase Sa and the 5K variant 121

28 Energetic analysis of the native state of RNase Sa and the 5K variant 122

LIST OF FIGURES

FIGURE Page

1 Ribbon diagram of RNase Sa (left) and RNase Sa3 (right). The figure was produced with Molscript (Kraulis 1991). 18

2 van der Waals attractive term versus measured ΔΔG value for Tyr to Phe mutation. A) Calculation using the AMBER force field B) Calculation using the CHARMM force field C) Calculation using the OPLS-AA force field. 33

3 Coulombic term versus measured ΔΔG value for Tyr to Phe mutation. A) Calculation using the AMBER force field B) Calculation using the CHARMM force field C) Calculation using the OPLS-AA force field. 34

4 Total electrostatic interactions divided into types. The interaction energy has been normalized for the size (number of atoms) of the protein. 50

5 Average electrostatic interactions between each residue and other atom groups in the protein. 52

6 Electrostatic interactions between individual groups and all other groups. A) Aliphatic Groups B) Aromatic Groups C) Polar Carbons D) Alpha Carbon E) Polar Groups. Groups in each panel have been sorted by the charge parameter of AMBER 99..... 53

7 Total Lennard-Jones attractive interactions divided into types. The interaction energy has been normalized for the size (number of atoms) of the protein. 59

8 Average Lennard-Jones attractive interaction between each residue and other atom groups in the protein. 60

9 Lennard-Jones attractive interactions between individual groups and all other groups. A) Aliphatic Groups B) Aromatic Groups C) Polar Carbons D) Alpha Carbon E) Polar Groups. Groups in each panel have been sorted by the van der Waals well depth parameter of AMBER 99. 62

FIGURE Page

10 Lennard-Jones repulsive interactions divided into types. The interaction energy has been normalized for the size (number of atoms) of the protein. 68

11 Average Lennard-Jones repulsive interaction between each residue and other atom groups in the protein. 69

12 Lennard-Jones repulsive interactions between individual groups and all other groups. A) Aliphatic Groups B) Aromatic Groups C) Polar Carbons D) Alpha Carbon E) Polar Groups. Groups in each panel have been sorted by the van der Waals well depth parameter of AMBER 99. 70

13 Atom Burial Classification..... 88

14 Mean volumes for each buried residue. A) Backbone volumes B) Side chain volumes. Error bars represent one standard deviation from the mean..... 88

15 Location of lysine substitutions in the 5K variant. The figure was produced with Molscript (Kraulis 1991)..... 104

16 Differences in the crystal structures. A) Variation in the backbone positions. Amino acids with a Cα RMSD >0.5 Å are represented by spheres centered at the Cα and proportional to the RMSD. B) Variation in the side chain positions. Amino acids with an average side chain RMSD>1.0 Å are represented by spheres centered at the Cβ and proportional to the average RMSD. The figure was produced with MolScript (Kraulis 1991). 110

17 Sample long range HNCO used to determine trans hydrogen bond J couplings in RNase Sa..... 116

INTRODUCTION

Proteins are essential for all living organisms. Proteins catalyze the chemical reactions necessary for life, as well as provide structure to the cell and regulate responses to the environment. To achieve these goals, proteins must fold into unique three-dimensional (3D) structures, yet the rules that determine the structure of a protein are still not fully understood.

Thermodynamic Hypothesis

The thermodynamic hypothesis states "the three-dimensional structure of a native protein in its normal physiological milieu (solvent, pH, ionic strength, presence of other components such as metal ions or prosthetic groups, temperature, etc) is the one in which the Gibbs free energy of the *whole system* is lowest; that is, that the native conformation is determined by the totality of interatomic interactions and hence by the amino acid sequence, in a *given environment*" (Anfinsen 1973). In short, the structure is defined by the sequence of the protein and the environment but not on the ribosome or other cellular machinery. Early work with hemoglobin and other proteins (Anson and Mirsky 1931; Anson 1945; Lumry and Eyring 1954) was able to show that denatured proteins could be renatured to have similar properties to the original proteins and the

---

This dissertation follows the style and format of *Protein Science*.

refolded proteins were indistinguishable from the original purified protein. Once it was shown that proteins could refold, there was interest in characterizing the thermodynamics of the transition between folded and denatured states. Several groups began to investigate the dependence of the unfolding on temperature, pressure and pH (Brandts and Lumry 1963; Acampora and Hermans 1967; Brandts et al. 1970; Zipp and Kauzmann 1973). Later, it became possible to use high-resolution calorimetry to show the thermodynamic reversibility of proteins (Pace 1975; Bolen and Santoro 1988; Santoro and Bolen 1988; Privalov 1989). Anfinsen was able to show that bovine pancreatic ribonuclease could return to the native state after the disulfide bonds were "scrambled" to a random distribution of 105 possible confirmations (Haber and Anfinsen 1962). For most proteins, the thermodynamic hypothesis has now been clearly established, and the simple answer to what determines the protein structure is the amino acid sequence. Interest has now turned to finding the rules that guide the protein to its 3D structure.

Historical View of Protein Folding

In 1936, Mirsky and Pauling had proposed that the structure in the polypeptide chain was held by hydrogen bonds between the backbone amides and carbonyls (Mirsky and Pauling 1936). Even before the first x-ray structure was solved, many of the regular secondary structural elements had been described solely on the basis of hydrogen bonding (Pauling and Corey 1951; Pauling et al. 1951).

In 1939, Bernal suggested the hydrophobic effect, the tendency for nonpolar groups to avoid interaction with the polar solvent, was the driving force for protein folding (Bernal 1939). It was not until Kauzmann's review in 1959 that the protein folding community began to view the hydrophobic effect as the dominant force in protein folding (Kauzmann 1959). Tanford measured the solubility of amino acids in water and argued that the hydrophobic effect would account for protein stability (Tanford 1962). In 1964, Brandts showed the large negative ΔCp of protein folding was also due to the hydrophobic effect (Brandts 1964).

Recently, hydrogen bonding has been regaining some of its prominence as a dominant force in protein stability. In fact, Pace and coworkers argued, based on mutational studies, that the contribution of hydrogen bonding and polar group burial was as significant as the hydrophobic effect (Myers and Pace 1996; Pace et al. 1996; Pace 2001).

Conformational Stability

Proteins in solution can occupy many different conformations, and all these conformations are in equilibrium. For many small globular proteins, however, it appears that only one of two major conformations exists and we can treat the equilibrium as:

1 $$N \leftrightarrow D$$

The native state (N) is the conformation seen in X-ray and NMR structures while the denatured state (D) is an ensemble of structures that is highly flexible and not well defined (Shortle 1996). The stability of the folded structure is thus defined as the free energy difference between the native and denatured states.

2 $$\Delta G_U = G_D - G_N$$

Under conditions where the protein is folded, ΔG_U would be positive. The folded state of globular proteins is generally 5-10 kcal/mol more stable, under physiological conditions, than the denatured state (Pace 1975). The stability can be determined from the relative populations of the native and denatured states using equation 3.

3 $$\Delta G_U = -RT \ln K_{eq}$$

K_{eq} is the equilibrium constant for unfolding, R is the gas constant (1.987 cal mol^{-1} K^{-1}) and T is the temperature in Kelvin. Experimentally this can be measured by perturbing the equilibrium constant using temperature, chemical denaturant (urea, GdnHCl), or other means.

Forces in Protein Folding

The stability of a protein can be defined in terms of entropy (ΔS) and enthalpy (ΔH).

4 $$\Delta G = \Delta H - T\Delta S$$

Entropy is a thermodynamic quantity related to the degree of randomness, or statistical probability of a system (Brady and Humiston 1986). For the unfolding reaction, there are two primary sources of entropy, the protein and the solvent. Enthalpy is the heat content of a system (Brady and Humiston 1986). Changes in enthalpy will come from changes in protein–protein, protein–solvent and solvent–solvent interactions.

The stability of the protein is a balance of several forces, some favoring the native state (hydrophobic effect, hydrogen bonding, packing) other favoring the denatured state (conformational entropy), and still others that can have a stabilizing or destabilizing effect (electrostatic interactions) depending on the circumstances. In the discussion below, the forces have been divided into those that are primarily entropic (hydrophobic effect and conformational entropy) and those that are primarily enthalpic (hydrogen bonds, electrostatic interactions and packing).

Hydrophobic Effect and Solvent Entropy

In 1939, Bernal came to the insightful conclusion that the tendency for the hydrophobic groups to be out of contact with water must play a significant role in protein folding (Bernal 1939). Kauzmann demonstrated that the hydrophobic effect is not an enthalpic, but rather an entropic change (Kauzmann 1959). Based on the hydrophobic effect, Kauzmann was able to understand two anomalies in protein folding: proteins unfold at low temperatures and at high pressure. These anomalies will be discussed in more detail later.

It is not energetically favorable for water, a polar medium, to solvate nonpolar compounds. The water molecules at the interface between the water and the nonpolar compound are oriented to minimize loss of hydrogen bonds. This results in loss of entropy in the system. By clustering nonpolar compounds, the number of restricted water molecules at the interface is reduced. Therefore, it is entropically favorable for nonpolar compounds to associate. This entropic penalty for water at the water-nonpolar interface gives rise to the hydrophobic effect. Since proteins contain a large number of nonpolar or hydrophobic groups and hydrophobic groups are generally more exposed to solvent in the denatured state than the native state, the hydrophobic effect favors protein folding.

The transfer of hydrophobic compounds to water has several important characteristics. There is a large positive change in heat capacity, ΔCp (Edsall 1935), the ratio $\Delta S/\Delta Cp=0.3$ at 25°C (Sturtevant 1977) and the temperature at which $\Delta S=0$ (T_s=110°C) and $\Delta H=0$ (T_H=22°C) (Baldwin 1986) are the same for many different hydrophobic compounds. Kauzmann's understanding of cold denaturation is therefore a consequence of the large positive ΔCp for the hydrophobic effect. Not surprisingly, the ΔCp for protein folding can be correlated with the amount of hydrophobic surface area that is exposed (Myers et al. 1995). At high and low temperatures, the exposure of non-polar surface reduces the entropy and enthalpy of the system. At low temperatures, the favorable ΔH for solvation of hydrophobic surface area is larger than the unfavorable ΔS, leading to cold denaturation (Tsai et al. 2002).

The second anomaly is that proteins unfold when the pressure is increased. Zipp and Kauzmann observed denaturation of Metmyoglobin as they increased the pressure to 85,000 PSI (Zipp and Kauzmann 1973) and Brandts et al denatured RNase A by increasing the pressure to 50,000 PSI (Brandts et al. 1970). Since the native state occupies a smaller volume than the denatured state, pressure should favor the native state if the protein were the only consideration. To understand the effects of pressure on protein unfolding, it is important to understand how pressure affects the structure of water. In ice, strong hydrogen bonding reduces the distance between interacting water molecules (Stillinger 1980). However, ice is less dense then liquid water because pockets are created in order to optimize the hydrogen-bonding network. At high pressure, water will form crystalline structures with two interpenetrating but unconnected hydrogen bond networks, filling the pockets (Eisenberg and Kauzmann 1969). At a nonpolar interface, the hydrogen-bonding network forms around the nonpolar compound, so the nonpolar compound occupies the cavity that would be formed in ice. Water at the nonpolar interface is more ordered, occupying a smaller volume than the bulk solvent. As a protein unfolds, more hydrophobic surface area is exposed which draws water from the bulk solvent into the denser nonpolar interface, decreasing the total volume of the system. Pressure therefore, favors the solvation of nonpolar compounds and causes the protein to unfold.

Conformational Entropy

Since the denatured state of a protein is highly flexible and poorly defined (Shortle 1996), there are a large number of thermodynamically equivalent conformations. The native state, on the other hand, consists of a small number of closely related conformations. Therefore, when the protein folds, there is a large reduction in the number of equivalent conformations which results in a loss of conformational entropy.

Kauzmann estimated that changes in conformational entropy favored the unfolded state by about 1.2 kcal/mol per residue at 25°C based on the average number of bonds per residue (Kauzmann 1954). Later, calorimetric studies by Privalov (Privalov 1979) were in reasonable agreement with Kauzmann's estimates. More recently, Spolar and Record used thermodynamic data to estimate a 1.7 kcal/mol per residue at 25°C contribution to conformational entropy (Spolar and Record 1994).

The contribution of the side chains to changes in conformational entropy can be useful in understanding helix propensities (Creamer and Rose 1992) and results of mutational studies in proteins (Myers and Pace 1996). Several different approaches have been used to estimate side chain conformational entropy. Doig and Sternberg compared results from different methods and concluded the cost of restricting side-chain motion was about 1.0 kcal/mol per residue or 0.5 kcal/mol per rotamer at 25°C (Doig and Sternberg 1995).

The restriction of the backbone upon folding is the other major component of conformational entropy. D'Aquino et al generated energy profiles of dipeptides to determine the difference in backbone conformational entropy between different residues (D'Aquino et al. 1996). Their calculations were in good agreement with differences in backbone conformational entropy experimentally determined from differences in stability of two alanine->glycine mutations in the coiled-coil of GCN4. Combining the estimates of backbone (D'Aquino et al. 1996) and side-chain (Lee et al. 1994) conformational entropy is in good agreement with the total values from Spolar and Record (Spolar and Record 1994).

Disulfide bonds also contribute to protein stability. Reduction of the disulphide bonds in RNase A results in the formation of a highly unfolded polypeptide chain (Harrington and Sela 1959). Flory attributed the stabilizing effect of disulfide bonds to the reduction of conformational entropy in the unfolded state (Flory 1956). Studies of the removal of disulfide bonds by mutagenesis or chemical modification have shown the contribution to conformational entropy varies by loop size, n (Pace et al. 1988).

5 $$\Delta S_{conf} = -2.1 - \frac{3}{2} R \ln n$$

Hydrogen Bonds

In 1920, Latimer and Rodebush described the hydrogen bond as a "hydrogen nucleus held between 2 octets" (Latimer and Rodebush 1920). They were able to explain many

of the properties of water with this “weak ‘bond’”. In 1928, Pauling argued that the hydrogen bond in which a proton holds two atoms of high electron affinity together is an electrostatic force (including polarization) (Pauling 1928). The nature and the strength of hydrogen bonds vary greatly. Relatively long (2.7 to 3.0 Å) hydrogen bonds are primarily electrostatic and are relatively weak (1 to 3 kcal/mol) whereas, short hydrogen bonds (2.3 to 2.5 Å) become primarily covalent and are relatively strong (~25 kcal/mol) (Gilli and Gilli 2000; Harris et al. 2000). Estimates of H-bond energies in ice are about 5.5 kcal/mol (Stillinger 1980). Intramolecular hydrogen bonds in biological molecules are primarily of the electrostatic type with an average energy of 5 kcal/mol (McDonald and Thornton 1994).

Proteins form an average of 1.1 hydrogen bonds per residue (Stickle et al. 1992), between the electronegative atoms of nitrogen and oxygen, and these interactions form the basis for the regular secondary structures (alpha helix and beta sheet) found in proteins (Pauling and Corey 1951; Pauling et al. 1951). Upon unfolding, intramolecular hydrogen bonds in the native state are replaced by protein-water hydrogen bonds in the denatured state. The contribution of the hydrogen bond to the stability of the protein is therefore dependent on the difference in energy between the hydrogen bonds in the denatured state and the hydrogen bonds in the native state (Equation 6).

$$H_2O\cdots HN + CO\cdots HOH \leftrightarrow CO\cdots HN + H_2O\cdots HOH$$
$$\Delta G = \Delta G_{H_2O\cdots HOH} + \Delta G_{CO\cdots HN} - \Delta G_{H_2O\cdots HN} - \Delta G_{CO\cdots HOH} \quad (6)$$

Theoretical studies have suggested hydrogen bonds contribute little to protein stability (Honig 1994; Lazaridis et al. 1995; Sippl et al. 1996), but experimental results have shown the removal of a single hydrogen bond can decrease protein stability by about 1.5 kcal/mol (Pace et al. 1996). Experimental results show some hydrogen bonds are more important than other hydrogen bonds. For example, tyrosine 86 in RNase Sa forms hydrogen bonds to the N^{ϵ} of arginine 69 and with an O^{ϵ} of glutamic acid 54 and removal of these hydrogen bonds by mutation of tyrosine 86 to phenylalanine results in marginal decrease in stability of 0.3 kcal/mol (Pace et al. 2001). On the other hand, tyrosine 51 forms a hydrogen bond to an O^{ϵ} of glutamic acid 78 and removal of this hydrogen bond results in a 2.3 kcal/mol loss in stability. Studies of glutamine-aspartic acid interactions in peptides show charge-neutral hydrogen bonds to contribute 1.0 kcal/mol versus a contribution of 0.4 kcal/mol for neutral-neutral hydrogen bonds (Huyghues-Despointes et al. 1995). Studies of proteins from thermophilic bacteria have shown a strong correlation between thermostability and the number of charged-netural hydrogen bonds (Tanner et al. 1996).

Electrostatic Interactions

Altering the pH of the solution was one of the first known methods to reversibly denature proteins (Anson and Mirsky 1931; Anson 1945). Changes in salt concentration can also affect protein stability (Acampora and Hermans 1967). Proteins contain several titratable groups: aspartic acid and glutamic acid are acidic residues that have a negative charge above the pKa while arginine, lysine and histidine are basic residues, having a

positive charge below the pKa. The N and C termini are also titratable groups. The pH at which there is an equal number of positive and negative charges on the protein is the isoelectric pH, pI.

The pKas of the titratable groups can be determined by monitoring proton uptake or release as a function of pH. In 1924, Linderstrom-Lang was able to model this behavior by treating the protein as an impenetrable sphere with a charge uniformly distributed over the surface (Linderstrom-Lang 1924). As a more detailed understanding of protein structure emerged, Tanford and Kirkwood were able to place charges at discrete positions in the protein (Tanford and Kirkwood 1957). Further enhancements to this model were made by Shire et al when they added a solvent accessibility term (Shire et al. 1974). An alternative method using the Poisson-Boltzmann equation was pioneered by Gilson and Honig (Gilson and Honig 1988). The Finite Difference Poisson-Boltzmann (FDPB) and solvent-accessible Tanford-Kirkwood (SATK) methods are generally able to predict the pKas of titratable groups in proteins (Huyghues-Despointes et al. 2003; Laurents et al. 2003).

In the denatured state, the titratable groups have pKas that are close to the pKas in model compounds (Elcock 1999; Kazmirski et al. 2001). However, in the native state, the pKas of titratable groups can be perturbed significantly. One example is Asp 79 in RNase Sa where the native state pKa is elevated by 3.4 pH units (Laurents et al. 2003). In this case, the environment of aspartic acid 79 has a strong preference for the protonated, neutral form of aspartic acid. Another example is aspartic acid 76 in RNase T1, which has a pKa

that is depressed by 3.5 pH units (Giletto and Pace 1999). Here, there is a strong preference for the deprotonated, charged form of aspartic acid. In each of these cases, the pKa is shifted because the favored state lowers the free energy of the native protein (relative to the disfavored state) by an amount equivalent to the free energy cost of perturbing the pKa. Thus, the pH dependence of the stability can be modeled from the differences in the pKas between the native and denatured state (Yang and Honig 1993).

Understanding the role of electrostatics in protein stability requires us to understand why pKas are perturbed in the native state. There are three major environmental influences on the pKa. Carboxyl groups that are buried in strongly hydrophobic pockets (like aspartic acid 79 in RNase Sa) have elevated pKas. Groups buried in polar environments that form many good hydrogen bonds often have depressed pKas (like Asp 76 in RNase T1). The other charges in the protein also affect the pKa. Coulomb's Law (Equation 7) tells us that like charges repel, unlike charges attract and the strength of these interactions depends on the distance between the charges.

7 $$E = \frac{q_1 q_2}{D r^2}$$

If the overall effect of the other charges is favorable, the charged form will be preferred (elevating pKas for basic groups and depressing pKas for acidic groups). Generally, these Coulombic effects can be screened by salt. However, salt cannot screen interactions if the two charged groups are close, ≤4 Å (Barlow and Thornton 1983), and form an ion pair or salt bridge.

The effect of salt bridges on protein stability is still not well understood. Studies of proteins from thermophilic bacteria show an increased number of salt bridges in the more stable proteins (Perutz and Raidt 1975), suggesting they play a role in stabilizing proteins. Theoretical studies based on electrostatic calculations conclude salt bridges are destabilizing (Hendsch and Tidor 1994). Anderson et al analyzed the pKa shifts in T4 Lysozyme and found salt-bridges to stabilize the folded state by 3-5 kcal/mol (Anderson et al. 1990). Attempts to stabilize proteins by engineering salt-bridges have been unsuccessful (Sali et al. 1991; Sun et al. 1991).

van der Waals Interactions and Packing

van der Waals interactions are typically described using the Lennard-Jones 6-12 potential (see equation 8) (Lennard-Jones 1931).

8 $$E_{vdW} = \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}}$$

The R^{-6} term represents the attraction between two induced dipoles caused by the non-uniform distribution of negatively charged electrons around the positively charged nucleus (Lennard-Jones 1931; London 1937). The distribution around one nucleus produces a dipole and that temporary dipole induces an opposing dipole in a neighboring atom. The dipole in the second atom reinforces the first dipole. Since the electron distribution is constantly changing, the directions of the dipoles change, but the changes in the two atoms are coupled providing a constant favorable induced dipole-induced

dipole interaction. London also investigated the use of additional terms including R^{-8} and R^{-10} terms to represent higher multipoles. He concluded that the contribution of the R^{-10} term always seemed negligible, but the R^{-8} term could become significant for He and H atoms at small distances (London 1937). The R^{-12} term represents a repulsive force between two atoms when the electron clouds overlap. The overlap of the electron clouds diminishes the shielding between the two nuclei, causing an electrostatic repulsion. Unfortunately, the dependence of the repulsion on the distance is not simply R^{-12}. Both London and Lennard-Jones recognized the need for an exponential term to accurately describe the repulsion at all distances. Lennard-Jones wrote "Theoretical calculations … show that the repulsive field is more complicated than this and contains terms of the form $e^{-\alpha R}$, but falls off very rapidly with distance and can (in the case of helium at any rate) be represented, over the range which is most effective in atomic collisions, by a term of the type $\lambda_{(rep.)}R^{-n}$."

In the native state, the protein interior approaches the packing density of close packed spheres (Richards 1974; Chothia 1975). Because of this, van der Waals forces are stronger in the native state than in the denatured state. It is difficult to study van der Waals forces directly with experimental methods since any mutational studies would also cause changes in other forces. In spite of this, there is experimental and computational evidence to support the role of van der Waals interactions in protein stability. Studies of leucine to alanine mutations in T4 Lysozyme have shown a relationship between the size of the cavity formed and the loss of protein stability (Matthews 1995). They attributed the loss in protein stability to a 1.9 kcal/mol change

from the loss in hydrophobicity due to the leucine to alanine substitution and a 24 cal/mol/Å^3 change from the formation of the cavity. Several groups have studied the role of van der Waals forces computationally. In computational studies of a threonine to valine substitution in T4 Lysozyme, Dang et al found the van der Waals component to be the dominant determinant of the change in stability (Dang et al. 1990). Prevost et al found the non-bonded interactions in the folded state account for the stability change in an isoleucine to alanine mutant in barnase (Prevost et al. 1991) and Sneddon and Tobias argue the stability change is caused by the loss of favorable packing interactions caused by cavity formation in two isoleucine to valine mutants in RNase T1 (Sneddon and Tobias 1992). Sugita and Kitao show the van der Waals forces again contribute to the change in stability of a isoleucine to valine mutation in human Lysozyme (Sugita and Kitao 1998). Kono et al showed that cavity-filling mutations contribute about 2 to 3 kcal/mol per methylene group (Kono et al. 2000), and by using computational methods, they were able to show the main contribution to be from van der Waals interactions associated with the cavity.

RNase Sa and Variants

Ribonuclease Sa (RNase Sa) is a small extracellular ribonuclease produced by *Streptomyces aureofaciens* strain BMK (Bacova et al. 1971). This 96 amino acid protein has a molecular weight of 10575 Da. RNase Sa contains no lysine, methionine or tryptophan residues, but has a high aromatic content with 8 tyrosine and 3 phenylalanine

residues. There is a single disulphide bond from cysteine 7 to cysteine 96, essentially linking the N and C termini of the protein.

Expression and purification of RNase Sa from *E. coli* in rich media was described by Hebert et al (Hebert et al. 1997). RNase Sa has also been isotopically enriched by expression in minimal media (Laurents et al. 1999). At pH 7, the conformational stability of RNase Sa at 25 °C is 6.1 kcal/mol and the melting temperature, Tm, is 48.4°C (Pace et al. 1998). Crystal structures of the native enzyme are available at 1.2Å (1rgg (Sevcik et al. 1996)) and 1.0Å (1lni (Sevcik et al. 2002a)) resolution. The structure is shown in Figure 1. The chemical shift assignments (Laurents et al. 1999) and NMR solution structure (1c54 (Laurents et al. 2001)) are also available.

RNase Sa is an acidic protein with a pI = 3.5. In 2001, Shaw et al described a variant of RNase Sa where they had replaced five surface acidic residues with five lysine residues (Shaw et al. 2001). The 5K variant (D1K+D17K+D25K+E41K+E74K) increased the pI to 10.2. The alteration of the net charge of the protein had little effect on the pH of maximum stability, but did significantly alter the dependence of solubility on pH.

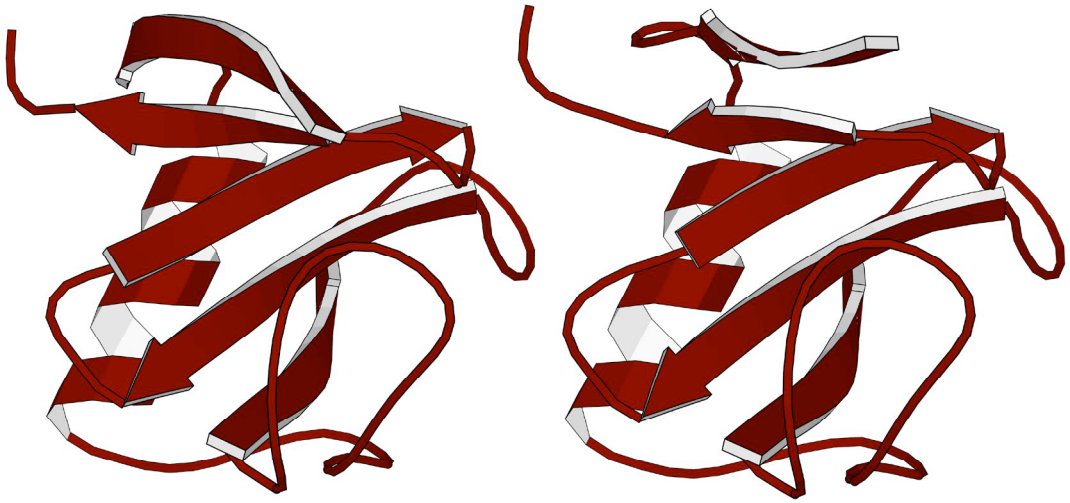


Figure 1: Ribbon diagram of RNase Sa (left) and RNase Sa3 (right). The figure was produced with Molscript (Kraulis 1991).

A closely related enzyme, RNase Sa3 is produced by an alternate strain of *Streptomyces aureofaciens*, strain CCM 3239 (Hebert et al. 1997). At pH 7, the conformational stability of RNase Sa3 at 25 °C is 5.6 kcal/mol and the melting temperature, Tm, is 47.2°C (Pace et al. 1998). The structure (see Figure 1) of RNase Sa3 (1mgr) was determined by Sevcik et al (Sevcik et al. 2002b). The sequences of RNase Sa and Sa3 are 69% identical and have similar structures (CαRMSD = 0.9Å).

Understanding the forces involved in protein folding and stability requires high-resolution structural data, mutational studies and computational analysis. High-resolution structural data is available for RNase Sa, the 5K variant, and RNase Sa3. Each of these proteins can be over-expressed in *E. coli* making it simple to create and analyze

new mutants. As detailed computational analyses are limited by the total number of atoms in the protein, the small size of these proteins makes a detailed computational analysis possible. The structural similarity coupled with the sequence differences between RNase Sa and RNase Sa3 allows for studies of the context dependence of mutations. While the structures are nearly the same, there are differences in the packing and specific interactions that provide insight into the role these forces play in protein stability. The reversal of the net charge in the 5K variant provides a unique opportunity to investigate the electrostatic nature of protein stability. Specific electrostatic interactions have not been altered, but the total electrostatic environment of the protein has completely changed. The result is a system that can be used to investigate the effect electrostatics on all aspects of protein stability.

In the following pages, four studies investigating various aspects of protein folding and structure are presented. In Tyrosine Hydroxyls, an investigation of the contribution of the tyrosine hydroxyl group to the overall stability of proteins is discussed. It had previously been shown that replacing a tyrosine generally destabilizes the protein, but there are large differences between the contribution of individual tyrosines (Pace et al. 2001). A correlation between the contribution of the tyrosine to stability and the van der Waals interactions of the tyrosine hydroxyl group is demonstrated. Electrostatic and van der Waals Interactions discusses an investigation of the energetic contribution of various atomic groups to the free energy of the native state. The van der Waals and electrostatic contributions of all the atomic groups in 2000 proteins are calculated and these data are analyzed to understand the average energetic contributions of individual amino acids and

their constituent atomic groups. The difference between polar groups, aliphatic carbons and aromatic carbons are also investigated. Packing of Buried Atoms discusses the dense packing in the protein interior. It has been proposed that hydrogen bonding increases the local packing in the protein. The local packing is analyzed for differences caused by hydrogen bonding as well as secondary structure. Structural Comparison of RNase Sa and 5K Variant shows a comparison of RNase Sa to the 5K variant. The crystal structures of both proteins are compared to look for differences in the backbone and the orientation of the side chains. Hydrogen bonding is also compared. In addition to the comparison of the crystal structure, NMR techniques are used to directly observe the NH···CO hydrogen bond in both proteins.

TYROSINE HYDROXYLS

Introduction

Tanford viewed the energetics of protein folding as a balance between hydrophobic interactions that promote protein folding and conformational entropy acting against protein folding (Tanford 1962). More recently, experimental evidence has suggested the contribution of hydrogen bonding and polar group burial to protein stability may be as large as the contribution of hydrophobic interactions (Fersht et al. 1985; Martensson et al. 1992; Serrano et al. 1992; Shirley et al. 1992; Chen et al. 1993; Byrne et al. 1995; Makhatadze and Privalov 1995; Myers and Pace 1996; Pace et al. 1996; Bhat et al. 1997; Koh et al. 1997; Yamagata et al. 1998; Takano et al. 1999; Albeck et al. 2000; Hamill et al. 2000; Pace et al. 2001). However, there is some resistance to the idea that hydrogen bonds contribute favorably to protein stability. When a protein is unfolded, all hydrogen bonding groups will form hydrogen bonds with water and many of these groups may not form all possible hydrogen bonds in the folded protein. Thus, the number of hydrogen bonds formed in the folded protein can never be greater than the number of hydrogen bonds in the unfolded protein. Some theoretical calculations also suggest that polar groups prefer to be fully solvated in water rather than buried in a protein even if they from intramolecular hydrogen bonds (Honig and Yang 1995). Honig and Yang concluded the burial of polar groups to be unfavorable based on the free energy of transfer of amino acids from water to non-polar solvents. Other work has found that the packing in the protein interior is greater than the packing in these solvents and

approaches the packing density of close packed spheres (Richards 1974). We believe that the increased packing in the protein interior can compensate for the penalty of moving the polar group from a polar solvent like water to the non-polar environment of the protein interior. In fact, studies of cavity formation suggest that the energetic loss of removing a non-polar group to form a cavity is from the loss in packing rather than the change in the hydrophobic burial (Kono et al. 2000).

It was recently proposed that the van der Waals interactions between buried polar groups contribute significantly to protein stability (Pace 2001). To gain further insight into the role polar groups play in protein stability, we need to understand the role of van der Waals forces in the burial of polar groups. Molecular force fields have been useful for understanding the changes caused by mutation (Fleischman and Brooks 1987; Tidor 1990; Sneddon and Tobias 1992). These force fields generally calculate the conformational energy of a system using a two body additive model with equations similar to Equation 9.

9 $$E_{Total} = \sum_{bonds} K_r(r - r_{eq})^2 + \sum_{angles} K_\theta(\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] + \sum_{i<j}[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\varepsilon R_{ij}}]$$

Although these molecular force fields do not include an explicit term for hydrogen bonds, Lifson, Hugler and Duaber demonstrated in 1979 that a 12-6-1 force field was able to accurately model hydrogen bonding interactions (Lifson et al. 1979). They concluded that the effects of the hydrogen bond (small van der Waals radii of the hydrogen, short hydrogen bond lengths and strong electrostatic interactions) all

originated from the electronegativity of the donor and acceptor atoms. They also compared their 12-6-1 force field to the MCMS force field (Momany et al. 1974; Dunfield et al. 1978) containing a 12-10 hydrogen bonding potential. Their results showed the additional parameters of the MCMS force field failed to improve the agreement with experimental results and in some cases was actually worse (Hagler et al. 1979).

For this analysis, I have chosen to focus on the non-bonded terms (Equation 10) since covalent bonds are not formed or broken during folding.

10 $$E_{NonBonded} = \sum_{i<j} [\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\varepsilon R_{ij}}]$$

These terms use the Lennard-Jones 12-6 potential to calculate van der Waals interactions and Coulombs Law to calculate coulombic interactions. I have chosen to break the van der Waals interactions into the component attractive and repulsive terms. Errors in the position of atoms will have a greater effect on the R^{-12} term because of its strong dependence on distance (Lomize et al. 2002). In hydrogen bonds, the strong electrostatic interaction brings the acceptor atom closer to the donor and hydrogen atoms. The inaccuracies of the R^{-12} term will therefore be greater at shorter distances.

As discussed above, Ribonuclease Sa (RNase Sa) is a small extracellular enzyme produced by *Steptomyces aureofaciens* (Bacova et al. 1971). The structure of RNase Sa has been studied by both x-ray crystallography (Sevcik et al. 1996; Sevcik et al. 2002a)

and NMR (Laurents et al. 2001). In 1998, Pace et al published studies of the thermodynamics of folding of RNases Sa, Sa2 and Sa3 (Pace et al. 1998). RNase Sa3 has 69% sequence identity and a similar structure (CαRMSD = 0.9 Å) to RNase Sa. Both enzymes contain eight tyrosines with seven at equivalent positions. Of these 16 tyrosines, four in RNase Sa and three in RNase Sa3 appear to form hydrogen bonds in the crystal structure.

Table 1 shows the extent of burial of each tyrosine as well as which positions are equivalent and which are hydrogen bonded. In 2001, Pace et al measured the conformational stability of all 16 tyrosine to phenylalanine mutants in RNase Sa and Sa3 (Pace et al. 2001). They observed that removing the polar hydroxyl group generally decreased protein stability, regardless of whether the tyrosine hydroxyl group forms intramolecular hydrogen bonds in the crystal structure or not. They also observed a wide range of effects on protein stability depending on which tyrosine was replaced. While there were some general trends to this variability, such as buried tyrosine that are hydrogen bonded generally contribute more to protein stability, no good quantitative measure investigated was able to predict the magnitude of the destabilization. Using a simple model to calculate the van der Waals and coulombic forces between the hydroxyl group and the rest of the protein, I want to see if I can find a correlation between the van der Waals attractive term and the change in stability caused by the substitution of tyrosine with phenylalanine.

Table 1: Burial of tyrosine hydroxyl

RNase Sa				RNase Sa3			
Site	Connolly Surface[1] (Å^2)	Solvent Accessible Surface[3] (Å^2)	% buried[3]	Site	Connolly Surface[1] (Å^2)	Solvent Accessible Surface[3] (Å^2)	% buried[3]
				11	19.5	22.5	57.5
30	16.3	18.1	65.8	33	26.6	32.1	39.4
49	30.8	41.5	21.6				
51	**7.7**	**8.7**	**83.5**	**54**	**1.5**	**1.6**	**97.1**
52	**0.0 (4.6[2])**	**0.0**	**100.0**	**55**	**0.0 (39.9[2])**	**0.0**	**100.0**
55	23.3	18.7	64.6	58	23.0	28.9	87.8
80	**6.6**	**6.4**	**88.0**	**83**	**2.4**	**6.6**	**87.5**
81	2.4	2.6	95.1	84	2.7	3.1	94.2
86	**2.2**	**7.4**	**85.9**	89	1.0	1.5	97.1

1 Calculated as in (Gerstein 1992)
2 Distance to closest surface atom (Å)
3 Calculated as in (Hebert et al. 1998)

Methods

Preparation of structures

The 1.2 Å (1rgg) (Sevcik et al. 1996) and 1.0 Å (1lni) (Sevcik et al. 2002a) x-ray crystal structures of RNase Sa were obtained from the PDB (www.rcsb.org/pdb/) (Berman et al. 2000). The structure of RNase Sa3 (1mgr) was obtained from Sevcik (Sevcik et al. 2002b). Existing hydrogen coordinates were removed from the 1.2 Å structure of RNase Sa so that all hydrogen atoms would be consistently positioned. Insight II (2000) from Accelrys was used to add the coordinates for the hydrogen atoms at pH 7.0. Insight II was then used to refine the location of the hydrogen atoms through energy minimization (1000 steps or the variance < 0.01) while the location of the heavy atoms remained fixed.

Table 2: Parameters for amino acids

		AMBER			CHARMM			OPLS		
Res	Atm	Charge	WellDepth	Radius	Charge	WellDepth	Radius	Charge	WellDepth	Radius
GLY	N	-0.4937	0.1700	1.8240	-0.4700	-0.2000	1.8500	-0.1400	0.1700	3.2500
	CA	-0.0332	0.1094	1.9080	-0.0200	-0.0200	2.2750	0.0800	0.0660	3.5000
	C	0.6731	0.0860	1.9080	0.5100	-0.1100	2.0000	0.5000	0.1050	3.7500
	O	-0.5854	0.2100	1.6612	-0.5100	-0.1200	1.7000	-0.5000	0.2100	2.9600
	H	0.3018	0.0157	0.6000	0.3100	-0.0460	0.2245	0.3000	0.0000	0.0000
	HA	0.0687	0.0157	1.3870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000
ALA	N	-0.4937	0.1700	1.8240	-0.4700	-0.2000	1.8500	-0.1400	0.1700	3.2500
	CA	0.0448	0.1094	1.9080	0.0700	-0.0200	2.2750	0.1400	0.0660	3.5000
	C	0.6731	0.0860	1.9080	0.5100	-0.1100	2.0000	0.5000	0.1050	3.7500
	O	-0.5854	0.2100	1.6612	-0.5100	-0.1200	1.7000	-0.5000	0.2100	2.9600
	CB	-0.0909	0.1094	1.9080	-0.2700	-0.0800	2.0600	-0.1800	0.0660	3.5000
	H	0.3018	0.0157	0.6000	0.3100	-0.0460	0.2245	0.3000	0.0000	0.0000
	HA	0.0228	0.0157	1.3870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000
	HB	0.0425	0.0157	1.4870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000
VAL	N	-0.4937	0.1700	1.8240	-0.4700	-0.2000	1.8500	-0.1400	0.1700	3.2500
	CA	-0.0530	0.1094	1.9080	0.0700	-0.0200	2.2750	0.1400	0.0660	3.5000
	C	0.6731	0.0860	1.9080	0.5100	-0.1100	2.0000	0.5000	0.1050	3.7500
	O	-0.5854	0.2100	1.6612	-0.5100	-0.1200	1.7000	-0.5000	0.2100	2.9600
	CB	0.3674	0.1094	1.9080	-0.0900	-0.0200	2.2750	-0.0600	0.0660	3.5000
	CG	-0.3584	0.1094	1.9080	-0.2700	-0.0800	2.0600	-0.1800	0.0660	3.5000
	H	0.3018	0.0157	0.6000	0.3100	-0.0460	0.2245	0.3000	0.0000	0.0000
	HA	0.0393	0.0157	1.3870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000
	HB	-0.0145	0.0157	1.4870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000
	HG	0.0803	0.0157	1.4870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000
LEU	N	-0.4937	0.1700	1.8240	-0.4700	-0.2000	1.8500	-0.1400	0.1700	3.2500
	CA	-0.0097	0.1094	1.9080	0.0700	-0.0200	2.2750	0.1400	0.0660	3.5000
	C	0.6731	0.0860	1.9080	0.5100	-0.1100	2.0000	0.5000	0.1050	3.7500
	O	-0.5854	0.2100	1.6612	-0.5100	-0.1200	1.7000	-0.5000	0.2100	2.9600
	CB	-0.1322	0.1094	1.9080	-0.1800	-0.0550	2.1750	-0.1200	0.0660	3.5000
	CG	0.4429	0.1094	1.9080	-0.0900	-0.0200	2.2750	-0.0600	0.0660	3.5000
	CD	-0.4312	0.1094	1.9080	-0.2700	-0.0800	2.0600	-0.1800	0.0660	3.5000
	H	0.3018	0.0157	0.6000	0.3100	-0.0460	0.2245	0.3000	0.0000	0.0000
	HA	0.0432	0.0157	1.3870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000
	HB	0.0317	0.0157	1.4870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000
	HG	-0.0662	0.0157	1.4870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000
	HD	0.1042	0.0157	1.4870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000
ILE	N	-0.4937	0.1700	1.8240	-0.4700	-0.2000	1.8500	-0.1400	0.1700	3.2500
	CA	-0.0257	0.1094	1.9080	0.0700	-0.0200	2.2750	0.1400	0.0660	3.5000
	C	0.6731	0.0860	1.9080	0.5100	-0.1100	2.0000	0.5000	0.1050	3.7500
	O	-0.5854	0.2100	1.6612	-0.5100	-0.1200	1.7000	-0.5000	0.2100	2.9600
	CB	0.0594	0.1094	1.9080	-0.0900	-0.0200	2.2750	-0.0600	0.0660	3.5000
	CG1	-0.0214	0.1094	1.9080	-0.1800	-0.0550	2.1750	-0.1200	0.0660	3.5000
	CG2	-0.3030	0.1094	1.9080	-0.2700	-0.0800	2.0600	-0.1800	0.0660	3.5000
	CD1	-0.0942	0.1094	1.9080	-0.2700	-0.0800	2.0600	-0.1800	0.0660	3.5000
	H	0.3018	0.0157	0.6000	0.3100	-0.0460	0.2245	0.3000	0.0000	0.0000
	HA	0.0640	0.0157	1.3870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000
	HB	0.0263	0.0157	1.4870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000
	HG1	0.0284	0.0157	1.4870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000
	HG2	0.0824	0.0157	1.4870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000
	HD1	0.0316	0.0157	1.4870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000
PRO	N	-0.2754	0.1700	1.8240	-0.2900	-0.2000	1.8500	-0.5000	0.1700	3.2500
	CA	-0.1503	0.1094	1.9080	0.0200	-0.1100	2.2750	0.0100	0.0660	3.5000
	C	0.6731	0.0860	1.9080	0.5100	-0.1100	2.0000	0.5000	0.1050	3.7500
	O	-0.5854	0.2100	1.6612	-0.5100	-0.1200	1.7000	-0.5000	0.2100	2.9600
	CB	0.0262	0.1094	1.9080	-0.1800	-0.0550	2.1750	-0.1200	0.0660	3.5000
	CG	0.0632	0.1094	1.9080	-0.1800	-0.0550	2.1750	-0.1200	0.0660	3.5000
	CD	0.0628	0.1094	1.9080	0.0000	-0.0550	2.1750	-0.0500	0.0660	3.5000
	HA	0.0758	0.0157	1.3870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000
	HB	0.0214	0.0157	1.4870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000
	HG	0.0063	0.0157	1.4870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000
	HD	0.0273	0.0157	1.3870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000

Table 2: Continued

		AMBER			CHARMM			OPLS		
Res	Atm	Charge	WellDepth	Radius	Charge	WellDepth	Radius	Charge	WellDepth	Radius
CYS	N	-0.4937	0.1700	1.8240	-0.4700	-0.2000	1.8500	-0.1400	0.1700	3.2500
	CA	0.0599	0.1094	1.9080	0.0700	-0.0200	2.2750	0.1400	0.0660	3.5000
	C	0.6731	0.0860	1.9080	0.5100	-0.1100	2.0000	0.5000	0.1050	3.7500
	O	-0.5854	0.2100	1.6612	-0.5100	-0.1200	1.7000	-0.5000	0.2100	2.9600
	CB	0.0027	0.1094	1.9080	-0.1000	-0.0550	2.1750	0.0975	0.0660	3.5000
	SG	-0.0893	0.2500	2.0000	-0.0800	-0.3800	1.9750	-0.2175	0.2500	3.5500
	H	0.3018	0.0157	0.6000	0.3100	-0.0460	0.2245	0.3000	0.0000	0.0000
	HA	0.0363	0.0157	1.3870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000
	HB	0.0473	0.0157	1.3870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000
SER	N	-0.4937	0.1700	1.8240	-0.4700	-0.2000	1.8500	-0.1400	0.1700	3.2500
	CA	0.0717	0.1094	1.9080	0.0700	-0.0200	2.2750	0.1400	0.0660	3.5000
	C	0.6731	0.0860	1.9080	0.5100	-0.1100	2.0000	0.5000	0.1050	3.7500
	O	-0.5854	0.2100	1.6612	-0.5100	-0.1200	1.7000	-0.5000	0.2100	2.9600
	CB	0.1345	0.1094	1.9080	0.0500	-0.0550	2.1750	0.1450	0.0660	3.5000
	OG	-0.5593	0.2104	1.7210	-0.6600	-0.1521	1.7700	-0.6830	0.1700	3.1200
	H	0.3018	0.0157	0.6000	0.3100	-0.0460	0.2245	0.3000	0.0000	0.0000
	HA	0.0164	0.0157	1.3870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000
	HB	0.0361	0.0157	1.3870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000
	HG	0.3687	0.0000	0.0000	0.4300	-0.0460	0.2245	0.4180	0.0000	0.0000
THR	N	-0.4937	0.1700	1.8240	-0.4700	-0.2000	1.8500	-0.1400	0.1700	3.2500
	CA	0.0702	0.1094	1.9080	0.0700	-0.0200	2.2750	0.1400	0.0660	3.5000
	C	0.6731	0.0860	1.9080	0.5100	-0.1100	2.0000	0.5000	0.1050	3.7500
	O	-0.5854	0.2100	1.6612	-0.5100	-0.1200	1.7000	-0.5000	0.2100	2.9600
	CB	0.4038	0.1094	1.9080	0.1400	-0.0200	2.2750	0.2050	0.0660	3.5000
	OG1	-0.6375	0.2104	1.7210	-0.6600	-0.1521	1.7700	-0.6830	0.1700	3.1200
	CG2	-0.2117	0.1094	1.9080	-0.2700	-0.0800	2.0600	-0.1800	0.0660	3.5000
	H	0.3018	0.0157	0.6000	0.3100	-0.0460	0.2245	0.3000	0.0000	0.0000
	HA	0.0164	0.0157	1.3870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000
	HB	-0.0909	0.0157	1.3870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000
	HG1	0.4051	0.0000	0.0000	0.4300	-0.0460	0.2245	0.4180	0.0000	0.0000
	HG2	0.0496	0.0157	1.4870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000
ASN	N	-0.4937	0.1700	1.8240	-0.4700	-0.2000	1.8500	-0.1400	0.1700	3.2500
	CA	0.0551	0.1094	1.9080	0.0700	-0.0200	2.2750	0.1400	0.0660	3.5000
	C	0.6731	0.0860	1.9080	0.5100	-0.1100	2.0000	0.5000	0.1050	3.7500
	O	-0.5854	0.2100	1.6612	-0.5100	-0.1200	1.7000	-0.5000	0.2100	2.9600
	CB	-0.1944	0.1094	1.9080	-0.1800	-0.0550	2.1750	-0.1200	0.0660	3.5000
	CG	0.6665	0.0860	1.9080	0.5500	-0.0700	2.0000	0.5000	0.1050	3.7500
	OD1	-0.5754	0.2100	1.6612	-0.5500	-0.1200	1.7000	-0.5000	0.2100	2.9600
	ND2	-0.7750	0.1700	1.8240	-0.6200	-0.2000	1.8500	-0.7600	0.1700	3.2500
	H	0.3018	0.0157	0.6000	0.3100	-0.0460	0.2245	0.3000	0.0000	0.0000
	HA	0.0410	0.0157	1.3870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000
	HB	0.0833	0.0157	1.4870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000
	HD2	0.3599	0.0157	0.6000	0.3000	-0.0460	0.2245	0.3800	0.0000	0.0000
GLN	N	-0.4937	0.1700	1.8240	-0.4700	-0.2000	1.8500	-0.1400	0.1700	3.2500
	CA	-0.0299	0.1094	1.9080	0.0700	-0.0200	2.2750	0.1400	0.0660	3.5000
	C	0.6731	0.0860	1.9080	0.5100	-0.1100	2.0000	0.5000	0.1050	3.7500
	O	-0.5854	0.2100	1.6612	-0.5100	-0.1200	1.7000	-0.5000	0.2100	2.9600
	CB	0.1411	0.1094	1.9080	-0.1800	-0.0550	2.1750	-0.1200	0.0660	3.5000
	CG	-0.2820	0.1094	1.9080	-0.1800	-0.0550	2.1750	-0.1200	0.0660	3.5000
	CD	0.7046	0.0860	1.9080	0.5500	-0.0700	2.0000	0.5000	0.1050	3.7500
	OE1	-0.6162	0.2100	1.6612	-0.5500	-0.1200	1.7000	-0.5000	0.2100	2.9600
	NE2	-0.7795	0.1700	1.8240	-0.6200	-0.2000	1.8500	-0.7600	0.1700	3.2500
	H	0.3018	0.0157	0.6000	0.3100	-0.0460	0.2245	0.3000	0.0000	0.0000
	HA	0.0381	0.0157	1.3870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000
	HB	-0.0021	0.0157	1.4870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000
	HG	0.1136	0.0157	1.4870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000
	HE2	0.3525	0.0157	0.6000	0.3000	-0.0460	0.2245	0.3800	0.0000	0.0000

Table 2: Continued

		AMBER			CHARMM			OPLS		
Res	Atm	Charge	WellDepth	Radius	Charge	WellDepth	Radius	Charge	WellDepth	Radius
ASP	N	-0.4937	0.1700	1.8240	-0.4700	-0.2000	1.8500	-0.1400	0.1700	3.2500
	CA	-0.1167	0.1094	1.9080	0.0700	-0.0200	2.2750	0.1400	0.0660	3.5000
	C	0.6731	0.0860	1.9080	0.5100	-0.1100	2.0000	0.5000	0.1050	3.7500
	O	-0.5854	0.2100	1.6612	-0.5100	-0.1200	1.7000	-0.5000	0.2100	2.9600
	CB	0.0021	0.0157	1.4870	-0.2800	-0.0550	2.1750	-0.2200	0.0660	3.5000
	CG	0.7672	0.0860	1.9080	0.6200	-0.0700	2.0000	0.7000	0.1050	3.7500
	OD	-0.7610	0.2100	1.6612	-0.7600	-0.1200	1.7000	-0.8000	0.2100	2.9600
	H	0.3018	0.0157	0.6000	0.3100	-0.0460	0.2245	0.3000	0.0000	0.0000
	HA	0.0656	0.0157	1.3870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000
	HB	-0.0210	0.0157	1.4870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000
GLU	N	-0.4937	0.1700	1.8240	-0.4700	-0.2000	1.8500	-0.1400	0.1700	3.2500
	CA	-0.2384	0.1094	1.9080	0.0700	-0.0200	2.2750	0.1400	0.0660	3.5000
	C	0.6731	0.0860	1.9080	0.5100	-0.1100	2.0000	0.5000	0.1050	3.7500
	O	-0.5854	0.2100	1.6612	-0.5100	-0.1200	1.7000	-0.5000	0.2100	2.9600
	CB	0.3053	0.1094	1.9080	-0.1800	-0.0550	2.1750	-0.1200	0.0660	3.5000
	CG	-0.1671	0.1094	1.9080	-0.2800	-0.0550	2.1750	-0.2200	0.0660	3.5000
	CD	0.6613	0.0860	1.9080	0.6200	-0.0700	2.0000	0.7000	0.1050	3.7500
	OE	-0.7362	0.2100	1.6612	-0.7600	-0.1200	1.7000	-0.8000	0.2100	2.9600
	H	0.3018	0.0157	0.6000	0.3100	-0.0460	0.2245	0.3000	0.0000	0.0000
	HA	0.0937	0.0157	1.3870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000
	HB	-0.0704	0.0157	1.4870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000
	HG	0.0313	0.0157	1.4870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000
LYS	N	-0.4937	0.1700	1.8240	-0.4700	-0.2000	1.8500	-0.1400	0.1700	3.2500
	CA	0.0343	0.1094	1.9080	0.0700	-0.0200	2.2750	0.1400	0.0660	3.5000
	C	0.6731	0.0860	1.9080	0.5100	-0.1100	2.0000	0.5000	0.1050	3.7500
	O	-0.5854	0.2100	1.6612	-0.5100	-0.1200	1.7000	-0.5000	0.2100	2.9600
	CB	-0.0196	0.1094	1.9080	-0.1800	-0.0550	2.1750	-0.1200	0.0660	3.5000
	CG	-0.0233	0.1094	1.9080	-0.1800	-0.0550	2.1750	-0.1200	0.0660	3.5000
	CD	-0.0574	0.1094	1.9080	-0.1800	-0.0550	2.1750	-0.1200	0.0660	3.5000
	CE	0.1461	0.1094	1.9080	0.2100	-0.0550	2.1750	0.1900	0.0660	3.5000
	NZ	-0.2135	0.1700	1.8240	-0.3000	-0.2000	1.8500	-0.3000	0.1700	3.2500
	H	0.0000	0.0157	0.6000	0.3100	-0.0460	0.2245	0.3000	0.0000	0.0000
	HA	0.0464	0.0157	1.3870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000
	HB	0.0143	0.0157	1.4870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000
	HG	0.0433	0.0157	1.4870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000
	HD	0.0602	0.0157	1.4870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000
	HE	0.0470	0.0157	1.1000	0.0500	-0.0220	1.3200	0.0600	0.0300	2.5000
	HZ	0.2872	0.0157	0.6000	0.3300	-0.0460	0.2245	0.3300	0.0000	0.0000
ARG	N	-0.4937	0.1700	1.8240	-0.4700	-0.2000	1.8500	-0.1400	0.1700	3.2500
	CA	0.0299	0.1094	1.9080	0.0700	-0.0200	2.2750	0.1400	0.0660	3.5000
	C	0.6731	0.0860	1.9080	0.5100	-0.1100	2.0000	0.5000	0.1050	3.7500
	O	-0.5854	0.2100	1.6612	-0.5100	-0.1200	1.7000	-0.5000	0.2100	2.9600
	CB	0.0040	0.1094	1.9080	-0.1800	-0.0550	2.1750	-0.1200	0.0660	3.5000
	CG	-0.0027	0.1094	1.9080	-0.1800	-0.0550	2.1750	-0.0500	0.0660	3.5000
	CD	0.1008	0.1094	1.9080	0.2000	-0.0550	2.1750	0.1900	0.0660	3.5000
	NE	-0.5111	0.1700	1.8240	-0.7000	-0.2000	1.8500	-0.7000	0.1700	3.2500
	CZ	0.8708	0.0860	1.9080	0.6400	-0.1100	2.0000	0.6400	0.0500	2.2500
	NH	-0.8348	0.1700	1.8240	-0.8000	-0.2000	1.8500	-0.8000	0.1700	3.2500
	H	0.3018	0.0157	0.6000	0.3100	-0.0460	0.2245	0.3000	0.0000	0.0000
	HA	0.0476	0.0157	1.3870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000
	HB	0.0257	0.0157	1.4870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000
	HG	0.0314	0.0157	1.4870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000
	HD	0.0575	0.0157	1.4870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000
	HE	0.3233	0.0157	0.6000	0.4400	-0.0460	0.2245	0.4400	0.0000	0.0000
	HH	0.4205	0.0157	0.6000	0.4600	-0.0460	0.2245	0.4600	0.0000	0.0000

Table 2: Continued

		AMBER			CHARMM			OPLS		
Res	Atm	Charge	WellDepth	Radius	Charge	WellDepth	Radius	Charge	WellDepth	Radius
PHE	N	-0.4937	0.1700	1.8240	-0.4700	-0.2000	1.8500	-0.1400	0.1700	3.2500
	CA	0.0416	0.1094	1.9080	0.0700	-0.0200	2.2750	0.1400	0.0660	3.5000
	C	0.6731	0.0860	1.9080	0.5100	-0.1100	2.0000	0.5000	0.1050	3.7500
	O	-0.5854	0.2100	1.6612	-0.5100	-0.1200	1.7000	-0.5000	0.2100	2.9600
	CB	-0.0021	0.1094	1.9080	-0.1800	-0.0550	2.1750	-0.0050	0.0660	3.5000
	CG	0.0293	0.0860	1.9080	0.0000	-0.0700	1.9924	-0.1150	0.0700	3.5500
	CD	-0.1050	0.0860	1.9080	-0.1150	-0.0700	1.9924	-0.1150	0.0700	3.5500
	CE	-0.1506	0.0860	1.9080	-0.1150	-0.0700	1.9924	-0.1150	0.0700	3.5500
	CZ	-0.0770	0.0860	1.9080	-0.1150	-0.0700	1.9924	-0.1150	0.0700	3.5500
	H	0.3018	0.0157	0.6000	0.3100	-0.0460	0.2245	0.3000	0.0000	0.0000
	HA	0.0184	0.0157	1.3870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000
	HB	0.0242	0.0157	1.4870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000
	HD	0.1123	0.0150	1.4590	0.1150	-0.0300	1.3582	0.1150	0.0300	2.4200
	HE	0.1161	0.0150	1.4590	0.1150	-0.0300	1.3582	0.1150	0.0300	2.4200
	HZ	0.1000	0.0150	1.4590	0.1150	-0.0300	1.3582	0.1150	0.0300	2.4200
TYR	N	-0.4937	0.1700	1.8240	-0.4700	-0.2000	1.8500	-0.1400	0.1700	3.2500
	CA	-0.0010	0.1094	1.9080	0.0700	-0.0200	2.2750	0.1400	0.0660	3.5000
	C	0.6731	0.0860	1.9080	0.5100	-0.1100	2.0000	0.5000	0.1050	3.7500
	O	-0.5854	0.2100	1.6612	-0.5100	-0.1200	1.7000	-0.5000	0.2100	2.9600
	CB	0.0211	0.1094	1.9080	-0.1800	-0.0550	2.1750	-0.0050	0.0660	3.5000
	CG	-0.0017	0.0860	1.9080	0.0000	-0.0700	1.9924	-0.1150	0.0700	3.5500
	CD	-0.1459	0.0860	1.9080	-0.1150	-0.0700	1.9924	-0.1150	0.0700	3.5500
	CE	-0.1848	0.0860	1.9080	-0.1150	-0.0700	1.9924	-0.1150	0.0700	3.5500
	CZ	0.2516	0.0860	1.9080	0.1100	-0.0700	1.9924	0.1500	0.0700	3.5500
	OH	-0.4757	0.2104	1.7210	-0.5400	-0.1521	1.7700	-0.5850	0.1700	3.0700
	H	0.3018	0.0157	0.6000	0.3100	-0.0460	0.2245	0.3000	0.0000	0.0000
	HA	0.0556	0.0157	1.3870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000
	HB	0.0225	0.0157	1.4870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000
	HD	0.1195	0.0150	1.4590	0.1150	-0.0300	1.3582	0.1150	0.0300	2.4200
	HE	0.1367	0.0150	1.4590	0.1150	-0.0300	1.3582	0.1150	0.0300	2.4200
	HH	0.3583	0.0000	0.0000	0.4300	-0.0460	0.2245	0.4350	0.0000	0.0000
TRP	N	-0.4937	0.1700	1.8240	-0.4700	-0.2000	1.8500	-0.1400	0.1700	3.2500
	CA	0.0233	0.1094	1.9080	0.0700	-0.0200	2.2750	0.1400	0.0660	3.5000
	C	0.6731	0.0860	1.9080	0.5100	-0.1100	2.0000	0.5000	0.1050	3.7500
	O	-0.5854	0.2100	1.6612	-0.5100	-0.1200	1.7000	-0.5000	0.2100	2.9600
	CB	-0.0168	0.1094	1.9080	-0.1800	-0.0550	2.1750	-0.1200	0.0660	3.5000
	CG	-0.0988	0.0860	1.9080	-0.0300	-0.0700	1.9924	0.0750	0.0700	3.5500
	CD1	-0.0808	0.0860	1.9080	0.0350	-0.0700	1.9924	-0.1150	0.0700	3.5500
	CD2	0.0768	0.0860	1.9080	-0.0200	-0.0900	1.8000	-0.0550	0.0700	3.5500
	NE1	-0.3542	0.1700	1.8240	-0.6100	-0.2000	1.8500	-0.5700	0.1700	3.2500
	CE2	0.1656	0.0860	1.9080	0.1300	-0.0900	1.8000	0.1300	0.0700	3.5500
	CE3	-0.2159	0.0860	1.9080	-0.1150	-0.0700	1.9924	-0.1150	0.0700	3.5500
	CZ2	-0.2279	0.0860	1.9080	-0.1150	-0.0700	1.9924	-0.1150	0.0700	3.5500
	CZ3	-0.1494	0.0860	1.9080	-0.1150	-0.0700	1.9924	-0.1150	0.0700	3.5500
	CH2	-0.1110	0.0860	1.9080	-0.1150	-0.0700	1.9924	-0.1150	0.0700	3.5500
	H	0.3018	0.0157	0.6000	0.3100	-0.0460	0.2245	0.3000	0.0000	0.0000
	HA	0.0376	0.0157	1.3870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000
	HB	0.0307	0.0157	1.4870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000
	HD1	0.1434	0.0150	1.4090	0.1150	-0.0300	1.3582	0.1150	0.0300	2.4200
	HE1	0.3403	0.0157	0.6000	0.3800	-0.0460	0.2245	0.4200	0.0000	0.0000
	HE3	-0.1561	0.0150	1.4590	0.1150	-0.0300	1.3582	0.1150	0.0300	2.4200
	HZ2	0.1382	0.0150	1.4590	0.1150	-0.0300	1.3582	0.1150	0.0300	2.4200
	HZ3	0.1087	0.0150	1.4590	0.1150	-0.0300	1.3582	0.1150	0.0300	2.4200
	HH2	0.1076	0.0150	1.4590	0.1150	-0.0300	1.3582	0.1150	0.0300	2.4200

Table 2: Continued

		AMBER			CHARMM			OPLS		
Res	Atm	Charge	WellDepth	Radius	Charge	WellDepth	Radius	Charge	WellDepth	Radius
HIS	N	-0.4937	0.1700	1.8240	-0.4700	-0.2000	1.8500	-0.1400	0.1700	3.2500
	CA	0.0367	0.1094	1.9080	0.0700	-0.0200	2.2750	0.1400	0.0660	3.5000
	C	0.6731	0.0860	1.9080	0.5100	-0.1100	2.0000	0.5000	0.1050	3.7500
	O	-0.5854	0.2100	1.6612	-0.5100	-0.1200	1.7000	-0.5000	0.2100	2.9600
	CB	-0.0413	0.1094	1.9080	-0.0413	0.0000	0.0000	-0.0050	0.0660	3.5000
	CG	0.2502	0.0860	1.9080	0.2502	0.0000	0.0000	-0.0150	0.0700	3.5500
	ND1	-0.5666	0.1700	1.8240	-0.5666	0.0000	0.0000	-0.4900	0.1700	3.2500
	CD2	-0.2281	0.0860	1.9080	-0.2281	0.0000	0.0000	0.0150	0.0700	3.5500
	CE1	0.2311	0.0860	1.9080	0.2311	0.0000	0.0000	0.2950	0.0700	3.5500
	NE2	-0.2110	0.1700	1.8240	-0.2110	0.0000	0.0000	-0.5700	0.1700	3.2500
	H	0.3018	0.0157	0.6000	0.3100	-0.0460	0.2245	0.3000	0.0000	0.0000
	HA	0.0217	0.0157	1.3870	0.0900	-0.0220	1.3200	0.0600	0.0300	2.5000
	1HB	0.0476	0.0157	1.4870	0.0476	0.0000	0.0000	0.0600	0.0300	2.5000
	HD2	0.1574	0.0150	1.4090	0.1574	0.0000	0.0000	-0.1150	0.0700	3.5500
	HE1	0.0679	0.0150	1.3590	0.0679	0.0000	0.0000	-0.1150	0.0700	3.5500
	HE2	0.2901	0.0157	0.6000	0.2901	0.0000	0.0000	0.4200	0.0000	0.0000

Calculation

The coulombic and Van der Waals contribution of the tyrosine hydroxyl was calculated using the CHARMM22 (Mackerell et al. 1998), AMBER (parm99) (Wang et al. 2000) and OPLS-AA (Jorgensen et al. 1996) parameter sets (see Table 2). The interactions were calculated between the Tyr Oη or the Tyr Hη and all other atoms separated from them by more than three bonds. The hydroxyl oxygen and hydrogen interactions were then added together for each of the van der Waals repulsive, van der Waals attractive, and the coulombic terms.

Results

van der Waals and Electrostatics

RNase Sa and RNase Sa3 each contain 8 tyrosine residues. In earlier work, each of the tyrosine residues was replaced with phenylalanine and the change in stability measured (Pace et al. 2001). To gain a better understanding of these results, I calculate the coulombic term and the van der Waals attractive and repulsive terms for the tyrosine hydroxyl group using the AMBER (Wang et al. 2000), OPLS (Jorgensen et al. 1996) and CHARMM (Mackerell et al. 1998) force fields. Table 3 shows a comparison between the measured ΔΔG for each mutant and the calculated electrostatic, attractive and repulsive van der Waals forces for the tyrosine hydroxyl group. I can see that hydrogen bonded tyrosine hydroxyl groups generally have a more favorable van der Waals attractive than the non hydrogen bonded tyrosine hydroxyl groups. I see a good correlation (approximately 0.7) between the measured ΔΔGs and the calculated van der Waals attractive term for all three force fields. The measured ΔΔG has been plotted against the calculated van der Waals attractive term in Figure 2. Table 4 shows the correlation coefficients for each calculated term versus ΔΔG.

Table 3: Calculation of total contributions of tyrosine hydroxyl group from molecule A of 1lni

			AMBER(parm99)[2]			OPLS-AA[3]			CHARMM22[4]		
Protein	Site	$\Delta\Delta G$[1]	LJA	LJR	Coulombic	LJA	LJR	Coulombic	LJA	LJR	Coulombic
RNase Sa	30	0.4	-3.1	0.9	23.6	-3.0	1.0	-77.5	-1.5	1.0	24.7
	49	-0.2	-1.9	0.6	26.9	-1.7	0.6	-52.4	-0.9	0.7	26.7
	51	**-2.3**	**-6.7**	**8.1**	**22.2**	**-6.2**	**7.5**	**-67.8**	**-3.2**	**8.3**	**15.9**
	52	**-3.6**	**-8.8**	**7.4**	**30.4**	**-8.1**	**7.0**	**-80.1**	**-4.4**	**8.0**	**29.6**
	55	-0.6	-3.3	0.8	36.3	-3.1	0.8	-74.2	-1.3	0.7	33.5
	80	**-1.5**	**-6.0**	**5.9**	**21.6**	**-5.6**	**5.5**	**-89.7**	**-2.9**	**6.0**	**16.1**
	81	-1.2	-6.1	3.7	24.9	-5.4	3.2	-84.3	-2.9	3.9	26.2
	86	**-0.3**	**-8.7**	**11.6**	**0.7**	**-7.3**	**9.9**	**-109.3**	**-4.2**	**11.6**	**-5.7**
RNase Sa3	11	-0.6	-3.2	1.2	31.1	-3.1	1.2	-48.6	-0.7	0.4	28.2
	33	0.5	-2.3	0.5	12.2	-2.1	0.5	-93.1	-1.1	0.6	7.7
	54	**-2.6**	**-7.1**	**8.7**	**4.4**	**-7.0**	**8.4**	**-83.1**	**-3.1**	**8.7**	**-9.2**
	55	**-2.1**	**-8.6**	**7.5**	**20.6**	**-8.1**	**7.1**	**-95.5**	**-3.8**	**7.2**	**11.9**
	58	-0.7	-3.1	0.7	28.9	-2.9	0.7	-95.7	-1.2	0.7	25.1
	83	**-1.5**	**-5.3**	**4.0**	**13.4**	**-4.9**	**3.7**	**-102.7**	**-2.5**	**4.0**	**3.9**
	84	-1.0	-5.1	2.3	4.4	-4.5	2.0	-105.8	-2.5	2.4	2.4
	89	0.0	-6.4	3.6	14.9	-6.0	3.7	-111.4	-2.9	3.8	6.9

LJA=-$\Sigma(B_j/R_j^6)$; LJR=$\Sigma(B_j/R_j^{12})$; all units in kcal/mol; sites in bold form intramolecular hydrogen bonds.

[1] (Pace et al. 2001)

[2] (Wang et al. 2000)

[3] (Jorgensen et al. 1996)

[4] (Mackerell et al. 1998)

Since hydrogen bonds are primarily an electrostatic interaction (Mitchell 1990), it is important to look at the role electrostatics plays in the mutants. Unlike the van der Waals term, there is little difference between the electrostatic term for hydrogen-bonded groups and those that are not (see Table 3). I also see no correlation between the calculated electrostatic term and the measured ΔΔG. The measured ΔΔG has been plotted against the calculated electrostatic contribution of the tyrosine hydroxyl in Figure 3.

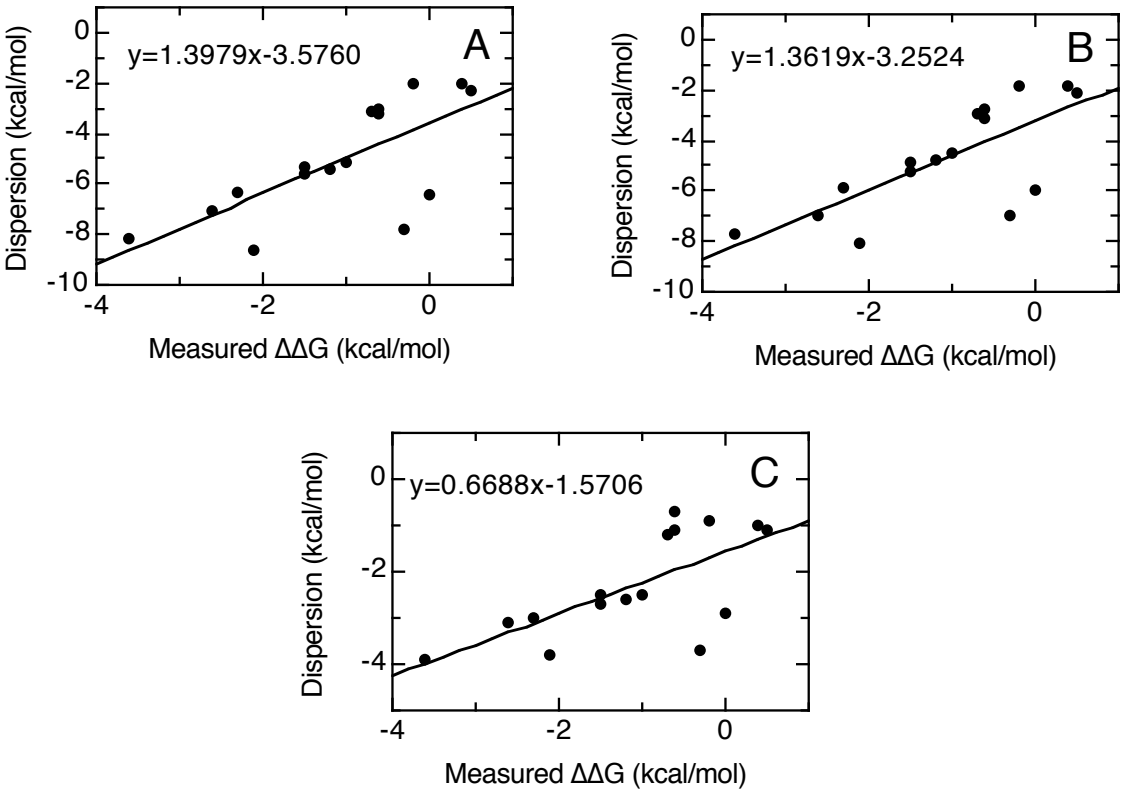


Figure 2: van der Waals attractive term versus measured ΔΔG value for Tyr to Phe mutation. A) Calculation using the AMBER force field B) Calculation using the CHARMM force field C) Calculation using the OPLS-AA force field.

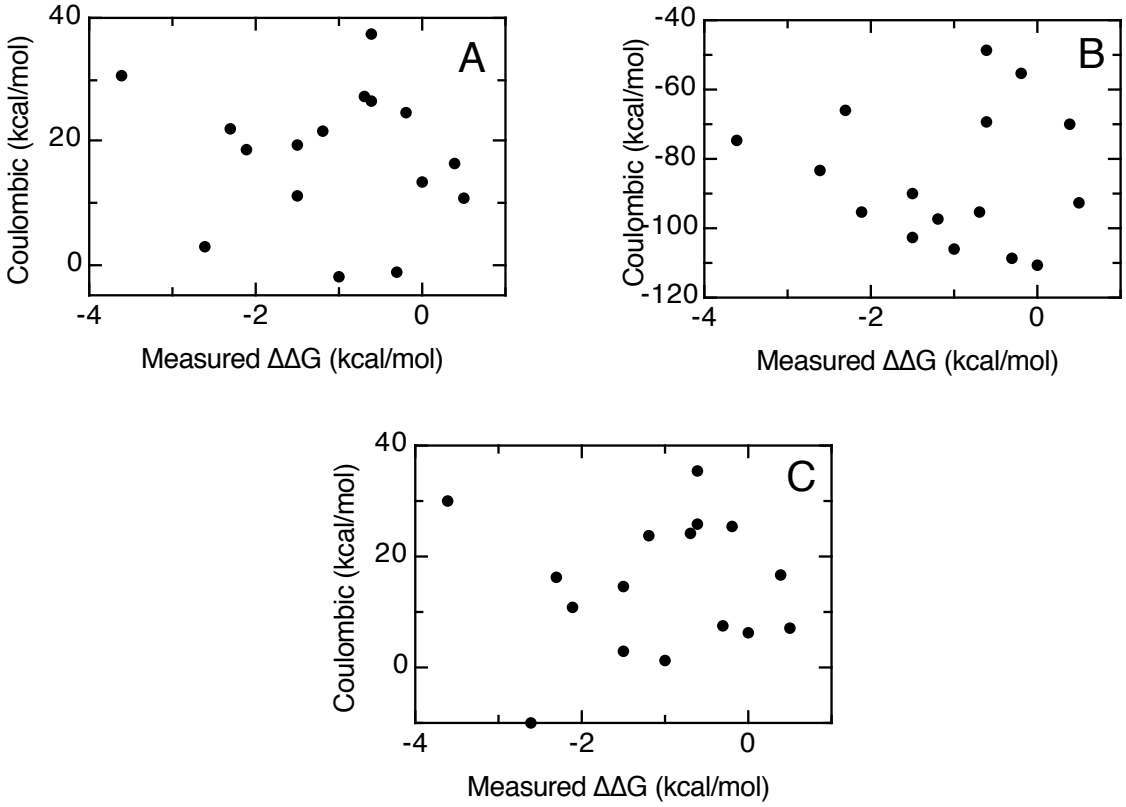


Figure 3: Coulombic term versus measured ΔΔG value for Tyr to Phe mutation. A) Calculation using the AMBER force field B) Calculation using the CHARMM force field C) Calculation using the OPLS-AA force field.

Table 4: Correlation coefficients between calculated terms and observed ΔΔG

	LJA	LJR	LJ Tot	ELS
AMBER	0.68	-0.61	-0.34	-0.05
OPLS	0.72	-0.64	-0.35	-0.06
CHARMM	0.66	-0.62	-0.57	0.04

$LJA=-\Sigma(B_j/R_j^6)$; $LJR=\Sigma(B_j/R_j^{12})$; LJTot=LJA+LJR; $ELS=\Sigma(q_iq_j/\varepsilon R_j)$

Table 5: Variation in AMBER calculations between different structures of RNase Sa

	1rgg A			1rgg B			1lni A			1lni B		
Site	LJA	LJR	Charge	LJA	LJR	Charge	LJA	LJR	Charge	LJA	LJR	Charge
30	-2.8	0.7	23.5	-2.0	0.4	16.2	-3.1	0.9	23.6	-2.1	0.5	19.0
49	-1.5	0.4	26.1	-2.0	0.6	24.8	-1.9	0.6	26.9	-2.4	0.7	24.1
51	**-6.3**	**7.1**	**20.6**	**-6.3**	**8.1**	**21.9**	**-6.7**	**8.1**	**22.2**	**-6.4**	**7.5**	**26.0**
52	**-8.6**	**7.4**	**30.2**	**-8.2**	**6.4**	**30.8**	**-8.8**	**7.4**	**30.4**	**-8.7**	**7.2**	**36.6**
55	-3.0	0.7	38.3	-3.0	0.6	37.2	-3.3	0.8	36.3	-3.1	0.7	38.1
80	**-5.5**	**6.0**	**22.1**	**-5.6**	**6.4**	**19.5**	**-6.0**	**5.9**	**21.6**	**-5.6**	**6.2**	**25.4**
81	-5.6	3.0	25.4	-5.4	2.5	21.8	-6.1	3.7	24.9	-5.9	3.3	26.0
86	**-15.9**	**103.6**	**14.3**	**-7.8**	**14.9**	**-1.1**	**-8.7**	**11.6**	**0.7**	**-7.8**	**9.2**	**2.5**

All units in kcal/mol, sites in bold form intramolecular hydrogen bonds.

To look at the variation in the calculations caused by changes in the crystal structure, I took advantage of the two high-resolution crystal structures available for RNase Sa (1rgg (Sevcik et al. 1996) and 1lni (Sevcik et al. 2002a)) and looked at the two molecules in each unit cell. Both structures were obtained under the same crystallization conditions. The crystallographic data for 1rgg were collected at room temperature and diffracted to 1.2 Å. The data for 1lni were collected at 100K and contain 1.0 Å diffractions. This gives us four independent molecules for the calculations and the results are shown in Table 5. I have focused on the AMBER force field, but the OPLS and CHARMM results are similar. With the exception of tyrosine 86, there is only a small difference in the calculations for the four molecules. Comparison of measured ΔΔG values with

calculated terms from all four structures result in similar correlation coefficients to those shown in Table 4.

Hydrogen Bonding

To look at the coulombic and van der Waals interactions of hydrogen-bonded groups, I compared the interactions with atoms within 3.5 Å of the tyrosine side-chain oxygen. Table 6 shows the coulombic and the van der Waals interactions using the AMBER force field. The groups that are hydrogen bonded have strong coulombic and van der Waals interactions with their hydrogen-bonding partner. The van der Waals interaction with the hydrogen bonding partner counts for a large portion of the overall van der Waals interactions for these groups, but the specific coulombic interactions seem to be masked by long range electrostatic interactions. Consider tyrosine 86; it is the only tyrosine in RNase Sa that forms two intramolecular hydrogen bonds: one as a donor and the other as an acceptor. The hydrogen bond to glutamic acid 54, (O-O distance of 2.55 Å) where tyrosine 86 acts as the donor, has a larger contribution to the van der Waals terms than the hydrogen bond with arginine 69 (O-N distance of 2.83 Å), where tyrosine 86 acts as an acceptor. While this may just be due to the geometries involved in this particular instance, it is known that tyrosine prefers to act as a donor rather than an acceptor (McDonald and Thornton 1994).

Table 6: Interaction of tyrosine hydroxyl group to atoms within 3.5 Å of the Oη

Site		Neighbor Residue		Neighbor Atom	Distance[1]	Coulombic Oη[2]	Coulombic Hη[2]	Coulombic Total[2]	van der Waals Attractive[2]	van der Waals Repulsive[2]
Tyr	30	Thr	64	O	3.23	28.70	-22.14	6.56	-0.56	0.37
		Arg	65	HA	3.29	-2.28	1.54	-0.74	-0.08	0.03
Tyr	49	Val	36	1HG2	3.12	-3.37	2.69	-0.67	-0.04	0.01
		Val	36	HB	3.36	0.68	-0.57	0.12	-0.09	0.03
Tyr	**51**	**Glu**	**78**	**OE2**	**2.55**	**45.61**	**-73.50**	**-27.89**	**-2.29**	**6.24**
		Ala	75	2HB	2.65	-2.53	1.90	-0.63	-0.36	0.57
		Ala	75	H	2.92	-16.30	13.90	-2.40	-0.03	0.00
		Gln	47	1HE2	3.22	-17.30	11.65	-5.65	-0.02	0.00
		Ala	75	3HB	3.42	-1.96	1.76	-0.20	-0.08	0.03
		Ala	75	CB	3.47	4.13	-3.36	0.77	-0.39	0.26
		Glu	78	CD	3.49	-29.90	31.08	1.18	-0.34	0.21
Tyr	**52**	Pro	45	1HD	2.62	-1.65	1.13	-0.52	-0.32	0.45
		Pro	**45**	**O**	**2.63**	**35.12**	**-41.79**	**-6.67**	**-1.89**	**4.25**
		Leu	44	1HB	2.73	-1.83	1.25	-0.58	-0.30	0.40
		Pro	45	2HB	3.06	-1.10	0.92	-0.18	-0.15	0.10
		Gln	47	2HG	3.08	-5.83	5.20	-0.63	-0.15	0.10
		Gln	47	1HB	3.09	-5.81	4.34	-1.47	-0.14	0.09
		Leu	44	2HB	3.31	-1.51	1.16	-0.35	-0.09	0.04
		Pro	45	N	3.39	12.85	-10.99	1.86	-0.50	0.33
		Pro	45	CD	3.4	-2.92	2.21	-0.71	-0.45	0.33
Tyr	55	Gly	34	2HA	2.85	-3.80	2.41	-1.39	-0.19	0.16
		His	53	2HB	3.35	3.35	2.11	5.46	-0.09	0.03
		His	53	1HB	3.46	-2.17	1.75	-0.42	-0.07	0.02
Tyr	**80**	**Glu**	**78**	**OE1**	**2.62**	**44.41**	**-53.39**	**-8.97**	**-1.95**	**4.54**
		Leu	91	1HD2	2.78	-5.92	3.89	-2.04	-0.27	0.32
		Glu	78	1HG	3.12	-1.59	1.25	-0.33	-0.14	0.08
		Pro	45	1HD	3.25	-1.33	0.93	-0.40	-0.09	0.03
Tyr	81	Arg	68	1HD	2.47	-3.68	3.41	-0.27	-0.55	1.32
		Pro	12	2HG	2.8	-0.36	0.28	-0.08	-0.26	0.29
		Arg	68	1HB	2.85	-1.42	1.32	-0.11	-0.23	0.24
		Arg	68	2HG	3.02	-1.64	1.20	-0.45	-0.17	0.12
		Pro	12	1HD	3.16	-1.36	1.05	-0.31	-0.10	0.05
		Arg	68	CD	3.31	-4.81	4.08	-0.74	-0.53	0.46
		Gly	83	2HA	3.34	-3.25	2.11	-1.14	-0.07	0.02
		Arg	68	CG	3.43	3.43	-0.10	3.33	-0.43	0.30
Tyr	**86**	**Arg**	**69**	**HE**	**1.86**	**-27.48**	**16.29**	**-11.19**	**-0.44**	**0.83**
		Glu	**54**	**OE2**	**2.55**	**45.60**	**-53.30**	**-7.69**	**-2.29**	**6.23**
		Arg	**69**	**NE**	**2.83**	**28.52**	**-18.95**	**9.57**	**-1.46**	**2.81**
		Arg	65	1HH1	3.03	-21.95	17.95	-4.01	-0.02	0.00
		Arg	69	1HG	3.18	-1.56	1.11	-0.45	-0.12	0.06
		Arg	69	2HH1	3.22	-20.63	12.50	-8.14	-0.02	0.00
		Arg	69	2HD	3.23	-2.81	2.30	-0.51	-0.11	0.05
		Glu	54	CD	3.37	-30.97	32.82	1.86	-0.42	0.32
		Glu	54	OE1	3.42	33.96	-34.26	-0.30	-0.39	0.18
		Arg	69	CD	3.47	-4.59	3.41	-1.18	-0.40	0.26

Items in bold are involved in an intramolecular hydrogen bond with the respective tyrosine

[1] Distances in Å

[2] kcal/mol

Discussion

The model I used in these calculations is deliberately simple. The primary difference between the wild-type protein and the variant with the phenylalanine substitution is the tyrosine hydroxyl. I ignore the possibility of structural rearrangement because it is generally small (Matthews 1995; Pace et al. 2001). I also ignore the differences in the solvation of phenylalanine and tyrosine in the unfolded protein. In the unfolded state, the tyrosine hydroxyl group should be exposed to solvent. Using a short peptide in a native-like conformation as a model for the unfolded state, the size of the tyrosine side chain would generally place the hydroxyl group in a solvent exposed position. Since the solvation difference between the phenylalanine and tyrosine will be primarily the result of the tyrosine hydroxyl group, these solvation differences in the unfolded state should be similar for all tyrosine to phenylalanine substitutions.

van der Waals

Figure 2 demonstrates a correlation between the van der Waals attractive term and the measured ΔΔG of the tyrosine to phenylalanine variants. The CHARMM force field produces a van der Waals attractive term of similar magnitude to the experimentally determined ΔΔG (slope=0.7) while AMBER and OPLS produce van der Waals attractive terms that are approximately twice as large (slope=1.4 and 1.3 respectively) as the values produced by CHARMM. However, the correlation between the attractive term and ΔΔG is better for the OPLS and AMBER force fields. Three points are significantly below the line and deserve more attention. Two of these variants (Y86F in RNase Sa and Y89F in

RNase Sa3) are at homologous positions. The hydroxyl of the tyrosine forms an intermolecular hydrogen bond to the substrate when it is bound in the active site. It is interesting that at this position an intramolecular hydrogen bonds is seen in the structure of RNase Sa, but not in the structure of RNase Sa3. Tyrosine 86 in RNase Sa is also the position where there is the most variation between the structures in Table 5. These sites may be involved in a dynamic process in solution that allows an interaction with substrate that is not observed in the substrate-free crystal structure, thus the contribution to the free energy would be smaller than expected from the crystallographic data. Looking at the structural differences between the substrate free from (1rgg) and the substrate bound form (1gmp) (Sevcik et al. 1993), glutamic acid 54, which makes a hydrogen bond to tyrosine 86 in 1rgg, moves away from tyrosine 86 in 1gmp to make room for the substrate. The other variant of interest is tyrosine 55 in RNase Sa3. This position is homologous to tyrosine 52 in RNase Sa and both have similar van der Waals contributions, but the measured ΔΔG for Y55F in RNase Sa3 (–2.6 kcal/mol) is considerably smaller than Y52F in RNase Sa (–3.6 kcal/mol). Most of the other homologous sites between RNase Sa and RNase Sa3 have similar ΔΔG values, so it would be expected that Y55F in RNase Sa3 to have a ΔΔG closer to –3.6 kcal/mol. This smaller observed value might indicate some rearrangement caused by the Y55F substitution in RNase Sa3 that would reduce the destabilizing effect of the Y55F substitution.

Electrostatics

While hydrogen bonds in proteins are primarily an electrostatic interaction between a partial positive charge on the hydrogen and the partial negative charge on the acceptor (McDonald and Thornton 1994), there is little difference in the coulombic term between tyrosine residues involved in hydrogen bonds and those that are not. The coulombic interactions are longer range than van der Waals interactions. This increases the contribution of distant atoms to the overall coulombic term, masking the effect of the local environment and hiding the contribution of the hydrogen bond. Looking specifically at the atoms involved in the hydrogen bond in Table 3, there is a large electrostatic contribution. As expected, the hydrogen-bonded group has a net favorable interaction to the acceptor.

The failure of the coulombic term to account for differences in the tyrosine ΔΔGs illustrates the need for more complete electrostatic models. The models used here do not account for polarization or charge transfer effects, although there is an effort to add these effects to molecular force fields (Meng et al. 1994; Dixon and Kollman 1997; Cieplak et al. 2001). The point charge models also fail to account for dipole-induced dipole interactions. Placing the dipole of the O–H bond in the interior of the protein should induce atomic dipoles in the neighboring atoms. These interactions may be small, but since the overall stability of proteins is also small relative to magnitude of the dominant forces (Dill 1990), it should not be overlooked.

Hydrogen Bonds

There is a striking difference in the van der Waals attractive term between tyrosine residues that are hydrogen bonded and those that are not. The magnitude of the van der Waals term in the hydrogen bonded tyrosine residues is considerably larger than for those that do not participate in intramolecular hydrogen bonds. Looking at Table 6, two components of the increased van der Waals interactions can be seen. Hydrogen bond partners contribute significantly to the van der Waals terms, contributing between 20% and 40% of the overall van der Waals attractive term. Also there is an increased number of atoms within 3.5Å of the Oη of a tyrosine involved in intramolecular hydrogen bonds. Honig proposed that hydrogen bonds increase the stability of proteins by increased packing, and our results support that view (Honig 1999). Increased packing caused by hydrogen bonds may also explain why Kuntz observed higher densities around polar groups than hydrophobic groups in proteins (Kuntz 1972).

Conclusions

Using this simplified model, I have shown that the calculated van der Waals interactions of the tyrosine hydroxyl group correlates with the observed ΔΔG for the tyrosine to phenylalanine mutations. This indicates that the observed stability changes are dependent on packing interactions. In support of this, it can been seen that tyrosine side chains that contribute more to stability contain more atoms within 3.5 Å of the Oη and form hydrogen bonds with strong contributions to the calculated van der Waals term

from the interaction between the tyrosine hydroxyl group and the hydrogen bonding partner.

ELECTROSTATIC AND VAN DER WAALS INTERACTIONS

Introduction

When proteins fold, more than 50% of all the side chains are buried in the interior of the molecule (Lesser and Rose 1990) and it has been shown that the packing of the protein approaches that of close packed spheres (Richards 1974). The burial of atom groups in the protein interior has a significant effect on protein stability. The hydrophobic effect requires the removal of non-polar groups from the protein solvent interface, and this clearly stabilized proteins. The role of polar group burial is less understood. Honig and Yang concluded that it is energetically unfavorable for polar groups to be buried and excluded from solvent (Honig and Yang 1995). Pace has argued that the burial of polar groups may make a strong favorable contribution to protein stability, if the increased packing in the protein interior is taken into account (Pace 2001).

Molecular force fields have been useful for the understanding of the atomic forces involved in protein folding (Fleischman and Brooks 1987; Tidor 1990; Sneddon and Tobias 1992). The energy is generally calculated using a two body additive model similar to Equation 9.

9 $$E_{Total} = \sum_{bonds} K_r(r - r_{eq})^2 + \sum_{angles} K_\theta(\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] + \sum_{i<j}[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\varepsilon R_{ij}}]$$

As above, I have chosen to focus on the non-bonded terms (Equation 10) as covalent bonds do not changed during protein folding.

10 $$E_{NonBonded} = \sum_{i<j} [\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\varepsilon R_{ij}}]$$

11 $$A_{ij} = (E_{ij} R_{ij})^{12} \qquad B_{ij} = (E_{ij} R_{ij})^{6}$$

12 $$E_{ij} = \sqrt{E_i E_j} \qquad R_{ij} = R_i + R_j$$

The Lennard-Jones 12-6 potential is used to calculate the van der Waals interactions and Coulombs Law is used to calculate interactions between partial charges. There are three parameters for each atom involved in calculating Equations 10-12: (1) the partial charge (q_i), (2) the van der Waals radius (R_i), and (3) the well depth or strength of the van der Waals interaction (E_i), which is a function of the polarizability of the atom. Atoms such as nitrogen and oxygen have a greater well depth than carbon atoms and benefit greatly from increased packing and van der Waals contacts. As above, the Lennard-Jones 12-6 potential has been calculated as two separate terms, due to the errors associated with the R^{-12} term that were previously discussed.

In an effort to improve our understanding of how interactions between various groups in the protein contribute to protein stability, I calculated the non-bonded interactions for all atom groups in a non-redundant set of 911 protein crystal structures. Since crystal structures contain little information about the placement of hydrogen atoms, the errors

associated with the R^{-12} term are exacerbated by the placement of hydrogen atoms on the structures.

Methods

Protein Data Set

PISCES (Wang and Dunbrack 2002) was use to generate a non-redundant set of high-resolution protein structures. The structures were required to be X-ray structures with resolution better than 1.8 Å, be less then 50% identical and have between 80 and 1000 amino acids. TINKER 3.9 (Ponder 2001) was used to add the coordinates for the hydrogen atom to all structures. After eliminating structures containing gaps and missing atoms, 911 structures remained. Table 7 contains the pdb codes and chain identifiers for these protein structures.

Calculation of Forces

The electrostatic and van der Waals interactions between atom pairs were calculated using the AMBER (parm 99) parameter set (Wang et al. 2000). The interactions were calculated between any atom pair separated by more than three bonds. Atoms were classified according to Table 8.

Table 7: Protein data set (911 Structures)

119l	A	1clx	A	1eg2	A	1g5t	A	1hg8	A	1jbk	A	1kw3	B	1mxr	A	1ra9	A
16pk	A	1cmc	A	1ehd	A	1g60	A	1hh8	A	1jbo	A	1kw4	A	1n08	A	1rcf	A
19hc	A	1cnz	A	1ejd	A	1g61	A	1hj9	A	1jcl	A	1kwn	A	1n0w	A	1rie	A
1a12	A	1co6	A	1ek0	A	1g66	A	1hm6	A	1jcx	A	1ky3	A	1n13	B	1rro	A
1a1i	A	1cqm	A	1ekg	A	1g6a	A	1hm9	A	1jd0	A	1kyf	A	1n1j	A	1rtu	A
1a2p	A	1cqx	A	1ekq	A	1g6g	A	1hmt	A	1jdr	A	1kyh	A	1n1j	B	1sgp	E
1a3a	A	1cru	A	1el5	A	1g6h	A	1hqk	A	1je0	A	1kyp	A	1n3b	A	1sgt	A
1a4i	A	1cs6	A	1elk	A	1g6s	A	1hqs	A	1jf2	A	1kzq	A	1n3y	A	1sml	A
1a62	A	1csh	A	1emv	A	1g8e	A	1htr	B	1jf3	A	1l0g	A	1n45	A	1svy	A
1a6m	A	1ctj	A	1emv	B	1g8k	B	1htw	A	1jf8	A	1l2h	A	1n55	A	1swu	A
1a73	A	1ctq	A	1eok	A	1g8q	A	1hx0	A	1jfb	A	1l6r	A	1n57	A	1tca	A
1a7s	A	1cuo	A	1ep0	A	1g9z	A	1hxh	A	1jfu	A	1l6x	A	1n62	A	1tfe	A
1a8d	A	1cv8	A	1eqo	A	1ga6	A	1hxi	A	1jfx	A	1l7a	A	1n71	A	1thf	D
1a8e	A	1cvl	A	1erv	A	1gai	A	1hyp	A	1jh6	A	1l7m	A	1n83	A	1thm	A
1a8q	A	1cxq	A	1erz	A	1gbs	A	1hz4	A	1jhd	A	1l9x	A	1n8k	A	1thx	A
1a9x	B	1cy9	A	1es5	A	1gc0	A	1hzo	A	1jhf	A	1lam	A	1n9p	A	1tml	A
1afw	A	1cyd	A	1es9	A	1gca	A	1hzt	A	1jhj	A	1lbv	A	1nbc	A	1toa	A
1ag9	A	1cyo	A	1eu3	A	1gci	A	1i0d	A	1jid	A	1lc0	A	1nep	A	1tx4	A
1agi	A	1czp	A	1euv	A	1gco	A	1i12	A	1jig	A	1ld8	A	1ney	A	1uah	A
1agj	A	1d02	A	1euw	A	1gcy	A	1i1n	A	1jiw	I	1ld8	B	1nff	A	1uch	A
1ah7	A	1d0c	A	1evh	A	1gd0	A	1i1w	A	1jix	A	1ldg	A	1nh8	A	1udh	A
1ajs	A	1d0q	A	1ew0	A	1gd1	O	1i2m	A	1jjf	A	1lf2	A	1nko	A	1ugi	A
1ak0	A	1d1q	A	1ew4	A	1gde	A	1i2s	A	1jk3	A	1lf7	A	1nkp	A	1uro	A
1ako	A	1d2n	A	1ew6	A	1gdo	A	1i40	A	1jkv	A	1lfk	A	1nkp	B	1uxy	A
1aky	A	1d2s	A	1exm	A	1gdv	A	1i4j	A	1jkx	A	1lfp	A	1nkr	A	1vca	A
1al3	A	1d2v	A	1exr	A	1geg	A	1i4u	A	1jl0	A	1lfw	A	1nlb	H	1vfr	A
1amf	A	1d2v	C	1ey4	A	1gg6	B	1i5g	A	1jl1	A	1lj5	A	1nls	A	1vhh	A
1amm	A	1d3g	A	1eye	A	1gg6	C	1i5r	A	1jlj	A	1lj8	A	1nme	A	1vpt	A
1aop	A	1d3v	A	1eyh	A	1ghe	A	1i60	A	1jlt	A	1lj9	A	1nme	B	1vsr	A
1aqb	A	1d4o	A	1eyv	A	1ghp	A	1i6l	A	1jlv	A	1lk5	A	1noa	A	1wer	A
1aqu	A	1d4t	A	1ezg	A	1giq	A	1i71	A	1jm1	A	1lkk	A	1nox	A	1whi	A
1aqz	A	1d7p	M	1ezw	A	1gj7	B	1i7h	A	1jnd	A	1lko	A	1npk	A	1xgs	A
1arb	A	1d8w	A	1f0j	A	1gk8	A	1i88	A	1jnr	B	1llp	A	1ns5	A	1xnb	A
1ars	A	1dbf	A	1f1m	A	1gk8	I	1i8o	A	1jo0	A	1lm5	A	1nsc	A	1yac	A
1atl	A	1dbx	A	1f2t	B	1gkl	A	1i9s	A	1jp3	A	1lni	A	1nsz	A	1ycc	A
1atz	A	1dc1	A	1f32	A	1gkp	A	1i9z	A	1jq5	A	1lo7	A	1nth	A	1yna	A
1auo	A	1dci	A	1f3u	A	1gmu	A	1ia6	A	1jr8	A	1lok	A	1nwz	A	1zin	A
1axn	A	1dd9	A	1f3u	B	1gmx	A	1iab	A	1jrr	A	1lop	A	1nxu	A	256b	A
1ay7	B	1ddw	A	1f46	A	1gnu	A	1iby	A	1jtg	B	1lpl	A	1ny1	A	2acy	A
1ayx	A	1df7	A	1f5v	A	1gny	A	1ic6	A	1jvw	A	1lq9	A	1nyt	A	2apr	A
1azo	A	1dfm	A	1f60	A	1go2	A	1icr	A	1jx4	A	1lqp	A	1nza	A	2ayh	A
1b0b	A	1dfu	P	1f60	B	1gp0	A	1id0	A	1jx6	A	1lqt	A	1nzy	A	2baa	A
1b16	A	1dg6	A	1f6b	A	1gpi	A	1ida	A	1jya	A	1lri	A	1o08	A	2bbk	H
1b2p	A	1dgw	A	1f74	A	1gpp	A	1iej	A	1jye	A	1ls1	A	1o7n	B	2bbk	L
1b5e	A	1dgw	Y	1f7d	A	1gpq	A	1ifc	A	1jyh	A	1ltz	A	1o8b	A	2bc2	A
1b5f	A	1dhn	A	1f86	A	1gq8	A	1ifr	A	1jyk	A	1lv7	A	1oa2	A	2bop	A
1b5f	B	1dix	A	1f8e	A	1gqa	A	1ift	A	1jyr	A	1ly2	A	1oaa	A	2bvw	A
1b6a	A	1dj0	A	1f8m	A	1gs5	A	1iib	A	1jzg	A	1lyc	A	1oaf	A	2cpl	A
1b8o	A	1dk0	A	1f9v	A	1gso	A	1ijq	A	1k07	A	1lyq	A	1oal	A	2cth	A
1b8z	A	1dk8	A	1f9z	A	1gtz	A	1ijy	A	1k0i	A	1lyv	A	1obd	A	2cua	A
1b9o	A	1dlf	L	1fao	A	1gu2	A	1ikt	A	1k0m	A	1lzj	A	1ock	A	2cy3	A
1b9w	A	1dlj	A	1faz	A	1gu7	A	1im5	A	1k1e	A	1lzl	A	1onc	A	2dri	A
1bbh	A	1dlw	A	1fbn	A	1gud	A	1in4	A	1k20	A	1m07	A	1one	A	2end	A
1bd0	A	1dly	A	1fc6	A	1guq	A	1inn	A	1k2e	A	1m15	A	1ooe	A	2fcb	A
1bdo	A	1dmg	A	1fcq	A	1gvk	B	1io0	A	1k3y	A	1m1n	A	1opd	A	2fcr	A
1beb	A	1dmh	A	1fcy	A	1gvo	A	1io7	A	1k4g	A	1m2d	A	1or3	A	2gdm	A
1bfg	A	1dnl	A	1fec	A	1gvp	A	1ioo	A	1k4i	A	1m3k	A	1pdo	A	2hft	A

Table 7: Continued

1bg6	A	1dos	A	1fgy	A	1gvu	A	1iq4	A	1k4n	A	1m40	A	1pgt	A	2hlc	A
1bgf	A	1dow	A	1fh0	A	1gwm	A	1iq6	A	1k55	A	1m45	A	1pin	A	2hmz	A
1bio	A	1doz	A	1fh9	A	1gx1	A	1iqc	A	1k5n	A	1m4i	A	1plc	A	2hvm	A
1bj7	A	1dpt	A	1fi2	A	1gx3	A	1iqq	A	1k5n	B	1m4j	A	1pot	A	2lis	A
1bjw	A	1dqe	A	1fiu	A	1gxm	A	1iqz	A	1k6w	A	1m4l	A	1ppf	E	2ltn	A
1bk7	A	1dqg	A	1fj2	A	1gxu	A	1ird	A	1k77	A	1m55	A	1ppn	A	2mcm	A
1bkb	A	1dqi	A	1fjh	A	1gxy	A	1ird	B	1k7c	A	1m5e	A	1psr	A	2mhr	A
1bkf	A	1dqp	A	1fjj	A	1gy6	A	1ire	A	1k7i	A	1m6p	A	1pym	A	2nac	A
1bkr	A	1dqz	A	1fk5	A	1gyo	A	1ire	B	1k7j	A	1m7y	A	1qaz	A	2nlr	A
1bm8	A	1ds1	A	1fl0	A	1gzg	A	1is3	A	1k8u	A	1m9z	A	1qb7	A	2por	A
1bn7	A	1dug	A	1flm	A	1gzt	A	1isp	A	1k94	A	1mba	A	1qcx	A	2pth	A
1bn8	A	1dus	A	1flt	X	1h03	P	1it2	A	1kae	A	1mc2	A	1qd1	A	2pvb	A
1bqb	A	1dwk	A	1fm0	D	1h05	A	1itx	A	1kaf	A	1mdc	A	1qd9	A	2sak	A
1bqk	A	1dxe	A	1fm0	E	1h0h	B	1iu8	A	1kao	A	1me4	A	1qe3	A	2sic	I
1brt	A	1dy5	A	1fmc	A	1h2e	A	1iua	A	1kdj	A	1meo	A	1qft	A	2spc	A
1bs0	A	1dym	A	1fn8	A	1h2r	S	1iup	A	1kep	A	1mfg	A	1qgi	A	2tgi	A
1bsm	A	1dys	A	1fn9	A	1h32	A	1iv3	A	1kew	A	1mgq	A	1qh4	A	2tnf	A
1bxa	A	1dz3	A	1fnd	A	1h4g	A	1ix9	A	1kfw	A	1mgt	A	1qh5	A	2tps	A
1bxv	A	1dzk	A	1fo8	A	1h4r	A	1ixh	A	1kgc	D	1mh9	A	1qhq	A	2trx	A
1byi	A	1e0c	A	1fp2	A	1h4x	A	1iz5	A	1kgs	A	1mix	A	1qhv	A	3cao	A
1byq	A	1e0w	A	1fqt	A	1h5q	A	1iz7	A	1khi	A	1mjh	A	1qip	A	3chb	D
1c02	A	1e19	A	1fs5	A	1h6f	A	1izc	A	1khx	A	1mjn	A	1qj5	A	3cyr	A
1c0p	A	1e1a	A	1fs7	A	1h6h	A	1j09	A	1kjw	A	1mk0	A	1qjc	A	3eip	A
1c1d	A	1e29	A	1ft5	A	1h6l	A	1j1x	H	1kli	H	1mkk	A	1qjp	A	3ezm	A
1c1k	A	1e2w	A	1ftr	A	1h6t	A	1j3a	A	1km4	A	1ml4	A	1qkk	A	3grs	A
1c1l	A	1e30	A	1fvg	A	1h6u	A	1j58	A	1koe	A	1mla	A	1ql0	A	3hts	B
1c44	A	1e43	A	1fvu	B	1h70	A	1j5r	A	1koi	A	1mml	A	1ql3	A	3lzt	A
1c52	A	1e4c	P	1fw9	A	1h72	C	1j6n	A	1kol	A	1mn8	A	1qmq	A	3nul	A
1c5e	A	1e58	A	1fxl	A	1h75	A	1j71	A	1kpf	A	1mna	A	1qnf	A	3pvi	A
1c7k	A	1e5m	A	1fxo	A	1h7c	A	1j83	A	1kpt	A	1mol	A	1qnn	A	3sil	A
1cb0	A	1e5p	A	1fye	A	1h7n	A	1j8b	A	1kq3	A	1moo	A	1qnr	A	3std	A
1cbs	A	1e7l	A	1fzq	A	1h7z	A	1j8f	A	1kqf	B	1mop	A	1qo7	A	3vub	A
1ccw	A	1e7w	A	1g0o	A	1h8u	A	1j8r	A	1kqf	C	1moq	A	1qop	A	4eug	A
1ccw	B	1e87	A	1g12	A	1h99	A	1j8u	A	1kqp	A	1mqk	H	1qop	B	4pga	A
1ccz	A	1ea7	A	1g16	A	1h9o	A	1j96	A	1kqr	A	1mqv	A	1qq4	A	4uag	A
1cex	A	1eaj	A	1g1t	A	1hbn	B	1j98	A	1kqw	A	1mr3	F	1qq5	A	4ubp	A
1cg5	A	1eaq	A	1g2a	A	1hbn	C	1j9b	A	1kr4	A	1msk	A	1qqf	A	4ubp	B
1cg5	B	1ear	A	1g2o	A	1hd2	A	1j9q	A	1kr7	A	1mty	B	1qs1	A	5nul	A
1chd	A	1eax	A	1g2q	A	1hdi	A	1ja9	A	1krh	A	1mty	G	1qsg	A	5pal	A
1ci9	A	1eaz	A	1g2r	A	1hdk	A	1jak	A	1ks8	A	1mug	A	1qst	A	5rub	A
1cip	A	1eb6	A	1g3p	A	1hdo	A	1jay	A	1ksh	B	1mun	A	1qtn	A	6gsv	A
1cjc	A	1ecs	A	1g4i	A	1hfe	S	1jb3	A	1kuf	A	1muw	A	1qtn	B	7a3h	A
1cjw	A	1ed8	A	1g4y	B	1hfo	A	1jb9	A	1kv7	A	1mwp	A	1qto	A	7fd1	A
1cke	A	1edg	A	1g57	A	1hfu	A	1jbe	A	1kv8	A	1mxi	A	1qtw	A	8tln	E
1cl8	A	1eeo	A														

Four letter PDB code and chain identifier.

Table 8: Classification of atoms

Aliphatic (Al)	ALA_CB, ARG_CB, ARG_CG, ARG_CD, ASN_CB, ASP_CB, CYS_CB, CYS_SG, GLN_CB, GLN_CG, GLU_CB, GLU_CG, HIS_CB, ILE_CB, ILE_CG1, ILE_CG2, ILE_CD1, LEU_CB, LEU_CG, LEU_CD1, LEU_CD2, LYS_CB, LYS_CG, LYS_CD, LYS_CE, MET_CB, MET_CG, MET_SD, MET_CE, PHE_CB, PRO_CB, PRO_CG, PRO_CD, SER_CB, THR_CB, THR_CG2, TRP_CB, TYR_CB, VAL_CB, VAL_CG1, VAL_CG2
Aromatic (Ar)	HIS_CG, HIS_CD2, HIS_CE1, PHE_CG, PHE_CD1, PHE_CD2, PHE_CE1, PHE_CE2, PHE_CZ, TRP_CG, TRP_CD1, TRP_CE2, TRP_CD2, TRP_CE3, TPR_CZ2, TRP_CZ3, TRP_CH2, TYR_CG, TYR_CD1, TYR_CD2, TYR_CE1, TYR_CE2, TYR_CZ
Polar Carbon (Cp)	ARG_CZ, ASN_CG, ASP_CG, GLN_CD, GLU_CD, All CO
Polar (P)	ARG_NE, ARG_NH1, ARG_NH2, ASN_OD1, ASN_ND2, ASP_OD1, ASP_OD2, GLN_OE1, GLN_NE2, GLU_OE1, GLU_OE2, HIS_ND1, HIS_NE2, LYS_NZ, SER_OG, THR_OG1, TRP_NE1, TYR_OH, All O, All N
Cα Carbons (Ca)	All CA

Hydrogens are in the same class as the covalently bonded atom. All refers to that position in all amino acids.

Results and Discussion

Electrostatic Interactions

Figure 4 shows the total electrostatic interactions in proteins normalized by the size of the protein (number of atoms). The interactions have been classified by the two atoms involved. Interactions between two polar groups create a large unfavorable electrostatic interaction. This reaction is unfavorable is due to the distribution of charge in the protein. Heavy polar atoms (nitrogen and oxygen) are given a partial negative charge. The directly attached hydrogen atoms are given a partial positive charge. This is because nitrogen and oxygen are electronegative atoms and the electron of the hydrogen atom delocalizes to the heavy polar atom. The polar carbons (those in a resonance structure with polar groups) are assigned a partial positive charge. For neutral groups, the

combination of the positive charge on the hydrogen and the polar carbon neutralizes the partial negative charge on the heavy polar atoms. For negatively charged side chains (aspartic acid and glutamic acid), there are no hydrogen atoms to balance the charge leaving a formal negative charge on the side chain. For positively charged side chains (lysine and arginine), the charge on the heavy polar atom is too small to balance the charge on the polar carbon and the hydrogen atoms leaving a formal positive charge on the side chain. Because the polar carbon has a partial positive charge, there is a strong favorable interaction between polar carbon groups and the negative polar groups in the rest of the protein. The net effect when looking at the combination of the polar-polar interactions and the polar-polar carbon interactions is a favorable electrostatic interaction between the polar side chains. Looking at the combination of aromatic-polar and aromatic-polar carbon interactions, there is another strong favorable interaction.

Figure 5 shows the average electrostatic interactions of each residue. For most residues, there are not any significant electrostatic interactions. There are a few exceptions. Positively charged residues (lysine and arginine) have strong favorable interactions with other polar groups, which have a partial negative charge, and unfavorable interactions with the polar carbons, which have a partial positive charge. This creates a net favorable interaction with polar amino acids in the protein. There is also a weak unfavorable interaction with the aliphatic groups and an intermediate unfavorable interaction with the aromatic groups.

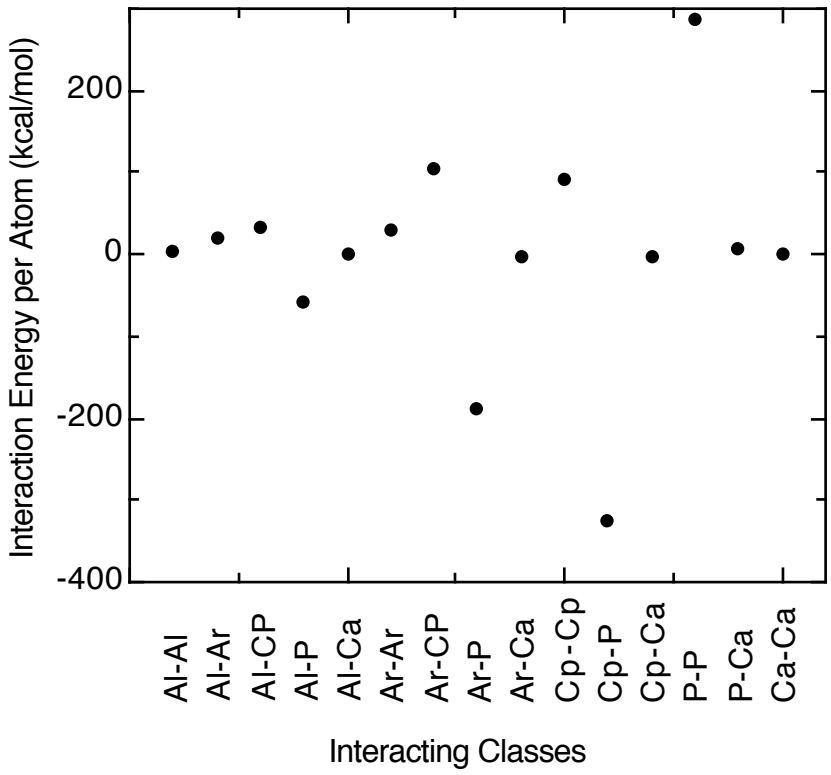


Figure 4: Total electrostatic interactions divided into types. The interaction energy has been normalized for the size (number of atoms) of the protein.

When a positive charge is positioned above or below the plane of the aromatic group, there is a favorable interaction between the cation and the π electrons of the aromatic group (Dougherty 1996). These cation-π interactions are common in proteins (Gallivan and Dougherty 1999). Gallivan and Dougherty found cation-π interactions involving arginine more common than those involving lysine and that tryptophan showed a strong preference to form these cation-π interactions. The data from Figure 5 show interactions between aromatic groups and the arginine and lysine amino acids as well as interactions

between tryptophan and polar groups. Interestingly, the data indicate these to be unfavorable interactions. Since the aromatic ring of the tryptophan contains a nitrogen atom, and for this work, the nitrogen was counted as a polar group, a larger portion of the negative charge is attributed to the nitrogen due to the electronegative character of the nitrogen. Therefore, from the point of view of the lysine or arginine, a portion of the favorable cation-π interaction is counted as a polar interaction. However, from the point of view of the tryptophan, the cation-π interaction includes contributions from polar and polar carbon groups. Along with the contribution of the cation-π interaction, there are also contributions with polar groups not involved in cation-π interactions, obscuring the true contribution of the cation-π interaction.

Glutamic acid and aspartic acid, the negatively charged amino acids, have unfavorable interactions with the negatively charged polar groups and favorable interactions with the positively charged polar carbons. The net effect is an unfavorable interaction with other polar amino acids. There is a weak favorable interaction with the aliphatic groups and an intermediate favorable interaction with the aromatic groups.

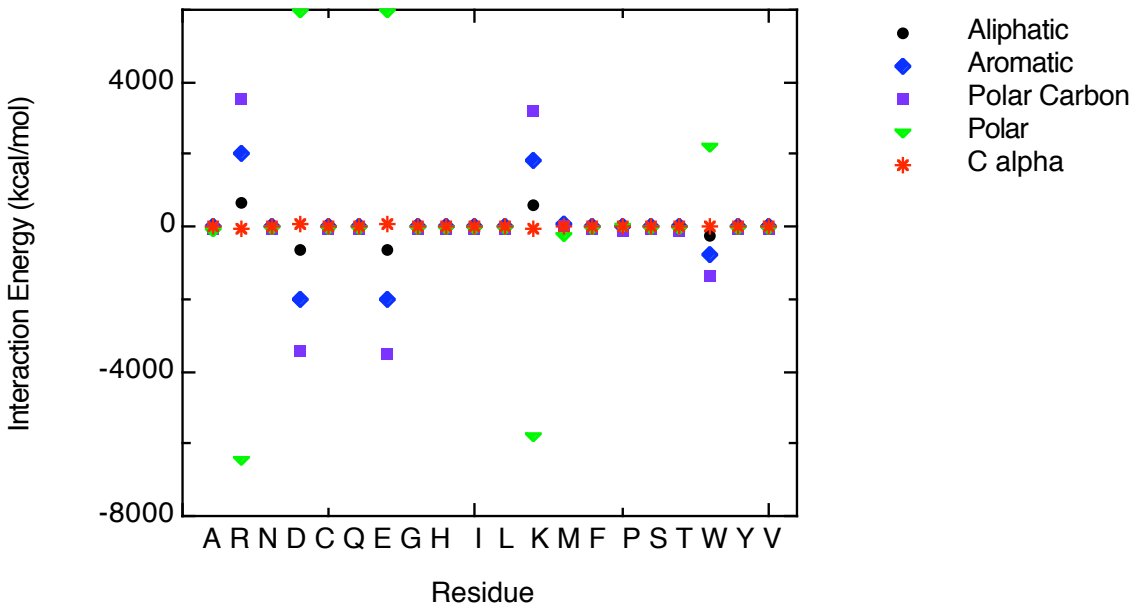


Figure 5: Average electrostatic interactions between each residue and other atom groups in the protein.

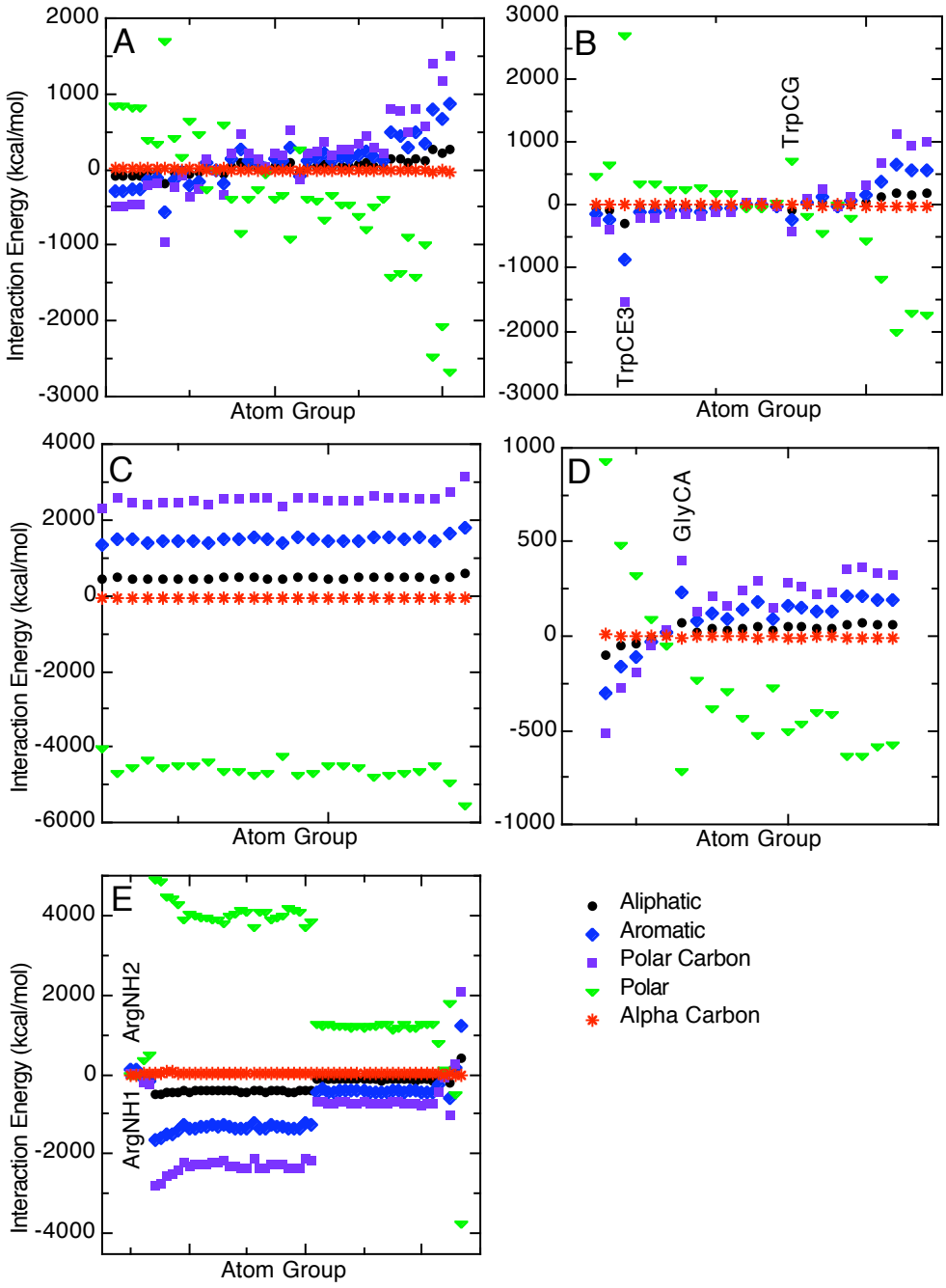


Figure 6: Electrostatic interactions between individual groups and all other groups. A) Aliphatic Groups B) Aromatic Groups C) Polar Carbons D) Alpha Carbon E) Polar Groups. Groups in each panel have been sorted by the charge parameter of AMBER 99.

Figure 6 shows the contribution of individual atom groups to the overall electrostatic interaction. Table 9 contains the data used to make the graphs. Since the electrostatic interactions are dependent upon the charge of the atom groups, the groups in each panel have been sorted by charge. There is a general trend for each set of atom groups. As the charge becomes more positive, there is a more favorable interaction with negatively charged polar groups and a less favorable interaction with positively charged groups (aliphatic, aromatic, polar and alpha carbons). A few atom groups do not fit this trend and deserve closer examination.

For the aromatic groups (panel B), there is a significant deviation for Cε3 and Cγ of tryptophan. Both of these atom groups are on the same side of the aromatic ring structure of tryptophan, indicating there is a strong preference for tryptophan to form electrostatic interactions on this side of the ring. In panel D, the Cα of glycine makes stronger interactions than would be expected. Without a side chain, the Cα of glycine is more accessible than the Cα of other amino acids. In panel E, at the far left are the Nη1 and Nη2 of arginine. The two groups have small interactions with all atom types. The arginine side chain is often exposed to solvent and therefore weakly interacts with the rest of the protein.

Table 9: Electrostatic interactions between the given atom group and class

Residue		Aliphatic	Aromatic	Polar Carbon	Polar	Alpha Carbon
ALA	C	480.7 ± 146.7	1513.0 ± 459.8	2605.0 ± 800.4	-4692.0 ± 1408.5	-48.0 ± 28.5
ALA	CA	48.6 ± 14.9	151.6 ± 46.7	264.3 ± 81.5	-466.4 ± 143.3	-4.8 ± 2.8
ALA	CB	26.2 ± 8.2	86.3 ± 25.6	149.7 ± 45.0	-264.8 ± 79.0	-2.6 ± 1.6
ALA	N	-128.1 ± 46.5	-433.1 ± 145.3	-742.6 ± 252.8	1250.0 ± 442.4	13.5 ± 8.3
ALA	O	-430.8 ± 128.0	-1337.0 ± 402.8	-2338.0 ± 700.0	4071.0 ± 1231.0	41.6 ± 24.9
ARG	C	476.3 ± 141.9	1534.0 ± 446.8	2545.0 ± 778.2	-4620.0 ± 1371.1	-51.3 ± 27.5
ARG	CA	54.9 ± 16.4	177.0 ± 51.8	294.9 ± 90.4	-526.4 ± 159.2	-5.9 ± 3.1
ARG	CB	38.4 ± 11.7	129.0 ± 37.0	213.2 ± 64.6	-381.2 ± 113.8	-4.3 ± 2.3
ARG	CD	143.8 ± 45.9	488.5 ± 144.8	797.8 ± 253.3	-1425.0 ± 446.4	-15.9 ± 8.6
ARG	CG	40.8 ± 12.7	139.2 ± 40.1	233.1 ± 70.1	-409.1 ± 123.5	-4.6 ± 2.4
ARG	CZ	576.1 ± 185.3	1804.0 ± 585.9	3156.0 ± 1024.3	-5572.0 ± 1807.0	-62.9 ± 34.6
ARG	N	-130.7 ± 43.1	-441.4 ± 133.9	-729.9 ± 231.9	1236.0 ± 407.7	14.6 ± 7.9
ARG	NE	-122.1 ± 40.0	-335.8 ± 126.1	-686.4 ± 220.5	1210.0 ± 388.5	13.5 ± 7.5
ARG	NH1	11.8 ± 5.0	108.8 ± 15.9	17.8 ± 28.6	-33.2 ± 50.7	-0.4 ± 0.8
ARG	NH2	7.3 ± 4.8	112.7 ± 14.6	10.7 ± 26.7	-21.8 ± 47.6	-0.4 ± 0.8
ARG	O	-423.2 ± 124.4	-1357.0 ± 393.5	-2290.0 ± 683.6	4014.0 ± 1203.7	44.7 ± 24.1
ASN	C	457.7 ± 144.5	1483.0 ± 458.9	2477.0 ± 796.9	-4548.0 ± 1404.0	-41.5 ± 27.6
ASN	CA	65.4 ± 20.8	211.4 ± 65.9	354.7 ± 114.6	-633.9 ± 201.8	-5.9 ± 3.9
ASN	CB	-18.8 ± 6.1	-41.7 ± 19.3	-92.2 ± 33.4	164.0 ± 58.9	1.7 ± 1.2
ASN	CG	442.0 ± 145.4	1418.0 ± 461.6	2435.0 ± 802.4	-4321.0 ± 1413.7	-40.7 ± 27.2
ASN	N	-120.6 ± 43.7	-426.0 ± 137.8	-688.4 ± 238.1	1215.0 ± 419.3	11.7 ± 7.9
ASN	NE2	-26.0 ± 12.3	-123.2 ± 39.2	-216.0 ± 67.6	377.3 ± 118.7	6.9 ± 2.4
ASN	O	-414.6 ± 127.0	-1310.0 ± 405.4	-2262.0 ± 701.6	3945.0 ± 1235.3	36.1 ± 24.1
ASN	OE1	-383.3 ± 126.1	-1222.0 ± 401.2	-2127.0 ± 695.6	3739.0 ± 1225.0	35.0 ± 23.4
ASP	C	459.1 ± 144.3	1459.0 ± 459.1	2464.0 ± 789.4	-4541.0 ± 1392.6	-44.5 ± 26.4
ASP	CA	-35.3 ± 11.0	-108.4 ± 35.0	-190.3 ± 60.2	320.3 ± 106.1	3.1 ± 2.0
ASP	CB	-26.8 ± 8.7	-86.7 ± 27.7	-146.6 ± 47.6	260.8 ± 83.9	2.6 ± 1.6
ASP	CG	508.5 ± 165.7	1647.0 ± 532.2	2769.0 ± 909.3	-4923.0 ± 1605.6	-49.5 ± 29.5
ASP	N	-128.2 ± 44.0	-419.3 ± 137.9	-683.1 ± 236.3	1215.0 ± 416.9	12.6 ± 7.6
ASP	O	-403.1 ± 126.5	-1290.0 ± 404.4	-2255.0 ± 692.7	3936.0 ± 1221.6	38.6 ± 23.1
ASP	OD1	-509.7 ± 164.6	-1635.0 ± 530.6	-2787.0 ± 903.9	4908.0 ± 1595.9	49.0 ± 29.3
ASP	OD2	-503.5 ± 165.2	-1615.0 ± 533.0	-2751.0 ± 907.7	4848.0 ± 1602.7	48.5 ± 29.0
CYS	C	460.5 ± 159.2	1446.0 ± 496.3	2484.0 ± 855.4	-4477.0 ± 1506.1	-40.7 ± 29.2
CYS	CA	66.8 ± 23.0	207.7 ± 71.7	361.3 ± 123.6	-638.0 ± 217.5	-5.8 ± 4.2
CYS	CB	67.9 ± 22.9	215.6 ± 71.4	373.6 ± 123.2	-661.6 ± 216.8	-6.0 ± 4.3
CYS	N	-127.5 ± 46.7	-417.4 ± 144.2	-715.5 ± 248.1	1201.0 ± 436.2	11.5 ± 8.3
CYS	O	-409.8 ± 139.6	-1285.0 ± 436.5	-2240.0 ± 751.0	3898.0 ± 1321.8	35.5 ± 25.6
CYS	SG	-62.4 ± 21.1	-197.9 ± 65.5	-343.7 ± 113.2	606.8 ± 199.3	5.5 ± 4.0
GLN	C	465.6 ± 141.1	1465.0 ± 447.6	2490.0 ± 777.7	-4497.0 ± 1369.8	-43.9 ± 26.6
GLN	CA	6.3 ± 2.0	18.4 ± 6.3	31.3 ± 11.0	-54.6 ± 19.5	-0.6 ± 0.5
GLN	CB	90.9 ± 28.8	297.3 ± 91.4	495.5 ± 159.1	-879.7 ± 280.2	-8.9 ± 5.4
GLN	CD	457.2 ± 149.8	1460.0 ± 475.7	2543.0 ± 827.7	-4475.0 ± 1457.6	-44.5 ± 27.1
GLN	CG	-35.0 ± 11.6	-99.3 ± 37.0	-184.6 ± 64.4	341.5 ± 113.3	3.5 ± 2.1
GLN	N	-137.7 ± 43.0	-421.9 ± 133.5	-714.3 ± 231.5	1202.0 ± 406.6	12.4 ± 7.6
GLN	NE2	-24.1 ± 16.0	-158.6 ± 50.8	-276.5 ± 88.1	484.9 ± 154.8	5.1 ± 2.9
GLN	O	-396.0 ± 123.8	-1296.0 ± 394.6	-2238.0 ± 684.7	3904.0 ± 1204.6	38.2 ± 23.2
GLN	OE1	-397.8 ± 131.7	-1273.0 ± 419.7	-2208.0 ± 728.2	3888.0 ± 1282.2	40.2 ± 23.7
GLU	C	451.8 ± 139.8	1419.0 ± 441.6	2422.0 ± 758.3	-4392.0 ± 1337.3	-48.3 ± 25.1
GLU	CA	-96.3 ± 30.0	-299.4 ± 95.0	-513.4 ± 163.2	925.4 ± 287.7	9.8 ± 5.3
GLU	CB	107.8 ± 34.2	336.7 ± 108.3	567.7 ± 185.8	-989.6 ± 327.7	-11.7 ± 6.1
GLU	CD	424.5 ± 138.1	1347.0 ± 441.3	2291.0 ± 751.6	-4041.0 ± 1326.4	-45.1 ± 23.5
GLU	CG	-66.7 ± 21.7	-216.6 ± 69.1	-360.6 ± 118.4	645.2 ± 208.8	7.3 ± 3.8
GLU	N	-134.8 ± 42.5	-408.6 ± 131.7	-695.6 ± 226.1	1173.0 ± 397.5	13.6 ± 7.2
GLU	O	-383.0 ± 122.7	-1257.0 ± 389.0	-2182.0 ± 667.0	3816.0 ± 1175.3	41.9 ± 22.0
GLU	OE1	-458.7 ± 154.2	-1494.0 ± 494.5	-2534.0 ± 839.5	4471.0 ± 1481.9	64.4 ± 26.2
GLU	OE2	-457.6 ± 154.4	-1486.0 ± 495.3	-2516.0 ± 840.3	4442.0 ± 1483.5	62.9 ± 26.1

Table 9: Continued

Residue		Aliphatic			Aromatic			Polar Carbon			Polar			Alpha Carbon		
GLY	C	475.6	±	152.5	1493.0	±	478.0	2568.0	±	832.6	-4630.0	±	1463.7	-45.9	±	28.9
GLY	CA	73.3	±	23.9	231.5	±	75.1	407.3	±	131.3	-715.1	±	230.7	-7.2	±	4.6
GLY	N	-131.2	±	46.3	-427.6	±	144.2	-732.6	±	250.8	1234.0	±	439.5	12.9	±	8.4
GLY	O	-420.4	±	133.2	-1320.0	±	420.1	-2306.0	±	730.7	4020.0	±	1283.4	39.8	±	25.2
HIS	C	473.1	±	153.8	1516.0	±	481.5	2562.0	±	844.8	-4667.0	±	1487.5	-49.0	±	27.6
HIS	CA	41.2	±	13.4	131.3	±	42.1	224.6	±	73.9	-398.7	±	130.0	-4.2	±	2.4
HIS	CB	38.0	±	12.4	125.6	±	39.0	214.0	±	68.7	-378.4	±	120.8	-4.0	±	2.2
HIS	CD2	-47.3	±	16.5	-154.3	±	51.9	-271.7	±	90.9	477.9	±	160.1	5.6	±	2.9
HIS	CE1	206.5	±	70.0	650.7	±	220.5	1135.0	±	386.8	-2000.0	±	681.5	-20.6	±	12.1
HIS	CG	174.1	±	57.8	556.8	±	181.7	961.3	±	318.9	-1705.0	±	561.8	-18.2	±	10.2
HIS	N	-127.2	±	46.3	-434.5	±	143.9	-723.9	±	251.5	1246.0	±	442.1	13.8	±	7.9
HIS	ND1	-392.7	±	131.5	-1250.0	±	414.1	-2188.0	±	725.9	3850.0	±	1278.6	40.9	±	23.1
HIS	NE2	52.0	±	19.2	168.0	±	60.8	291.5	±	106.7	-516.0	±	188.4	-5.4	±	3.2
HIS	O	-424.0	±	133.8	-1343.0	±	420.5	-2319.0	±	736.8	4058.0	±	1296.4	42.7	±	24.1
ILE	C	509.0	±	145.7	1534.0	±	455.3	2637.0	±	792.9	-4748.0	±	1394.5	-50.3	±	27.9
ILE	CA	30.0	±	8.4	89.3	±	26.2	152.0	±	45.7	-272.4	±	80.3	-3.0	±	1.7
ILE	CB	63.6	±	18.6	198.5	±	58.0	343.7	±	101.3	-608.4	±	178.2	-6.5	±	3.6
ILE	CD1	-2.2	±	0.4	1.1	±	1.2	1.9	±	2.2	-3.4	±	3.8	-0.2	±	0.1
ILE	CG1	28.0	±	7.7	82.6	±	24.0	146.9	±	42.0	-256.4	±	73.9	-2.7	±	1.5
ILE	CG2	-41.1	±	12.1	-130.8	±	37.8	-209.3	±	65.9	390.8	±	115.9	4.3	±	2.3
ILE	N	-147.9	±	44.2	-440.3	±	135.6	-755.1	±	235.8	1271.0	±	413.6	14.2	±	8.0
ILE	O	-431.3	±	127.2	-1359.0	±	399.2	-2368.0	±	693.5	4125.0	±	1219.3	43.7	±	24.3
LEU	C	466.9	±	143.8	1527.0	±	451.5	2618.0	±	785.2	-4718.0	±	1381.4	-50.6	±	27.0
LEU	CA	22.4	±	7.3	77.2	±	22.7	132.1	±	39.6	-235.7	±	69.6	-2.6	±	1.4
LEU	CB	-46.4	±	14.7	-155.9	±	46.3	-269.8	±	80.7	477.6	±	141.9	5.3	±	2.8
LEU	CD1	-76.1	±	25.6	-275.5	±	80.2	-485.3	±	140.2	850.5	±	246.3	8.8	±	4.8
LEU	CD2	-76.1	±	25.6	-275.1	±	80.2	-484.6	±	140.1	849.3	±	246.3	8.8	±	4.8
LEU	CG	276.8	±	81.2	875.2	±	254.5	1514.0	±	444.5	-2680.0	±	781.2	-28.9	±	15.2
LEU	N	-113.5	±	42.8	-439.4	±	133.0	-751.1	±	230.9	1265.0	±	405.2	14.3	±	7.7
LEU	O	-439.8	±	126.2	-1353.0	±	398.4	-2355.0	±	691.0	4103.0	±	1215.0	44.0	±	23.6
LYS	C	454.0	±	137.0	1397.0	±	428.6	2350.0	±	747.2	-4259.0	±	1317.3	-47.5	±	26.1
LYS	CA	54.6	±	16.4	166.7	±	51.5	282.2	±	89.9	-501.9	±	158.4	-5.7	±	3.1
LYS	CB	6.6	±	2.0	19.6	±	6.3	32.9	±	10.9	-58.7	±	19.3	-0.6	±	0.4
LYS	CD	39.8	±	12.7	132.7	±	39.8	213.9	±	69.7	-379.9	±	123.0	-3.9	±	2.4
LYS	CE	146.7	±	48.0	485.3	±	150.8	798.3	±	263.7	-1418.0	±	465.4	-15.5	±	8.8
LYS	CG	41.0	±	12.8	132.0	±	40.0	223.4	±	69.9	-393.1	±	123.3	-4.4	±	2.4
LYS	N	-125.2	±	42.9	-401.3	±	131.8	-672.5	±	228.4	1131.0	±	401.2	13.3	±	7.6
LYS	NZ	414.6	±	129.4	1239.0	±	408.0	2103.0	±	713.0	-3774.0	±	1259.8	-41.2	±	23.5
LYS	O	-401.4	±	121.4	-1238.0	±	380.0	-2121.0	±	660.6	3703.0	±	1163.7	41.3	±	22.9
MET	C	484.6	±	151.7	1531.0	±	473.6	2622.0	±	824.3	-4723.0	±	1450.5	-50.4	±	27.4
MET	CA	30.4	±	10.7	95.5	±	33.2	164.6	±	58.0	-291.9	±	102.2	-3.1	±	1.8
MET	CB	42.0	±	12.7	128.2	±	39.5	221.4	±	68.9	-391.9	±	121.2	-4.3	±	2.3
MET	CE	89.0	±	27.9	271.9	±	86.6	476.2	±	152.1	-836.3	±	266.8	-8.9	±	4.9
MET	CG	94.7	±	28.9	294.1	±	89.9	517.4	±	157.4	-906.4	±	276.3	-9.6	±	5.2
MET	N	-128.9	±	62.5	-417.0	±	196.9	-712.3	±	339.0	1198.0	±	588.6	13.3	±	9.3
MET	O	-425.2	±	131.5	-1355.0	±	411.3	-2353.0	±	717.5	4101.0	±	1259.8	43.7	±	23.9
MET	SD	-176.1	±	55.6	-552.7	±	173.1	-970.5	±	303.0	1702.0	±	531.9	18.1	±	9.9

Table 9: Continued

Residue		Aliphatic	Aromatic	Polar Carbon	Polar	Alpha Carbon
PHE	C	483.2 ± 147.2	1516.0 ± 465.2	2610.0 ± 809.4	-4704.0 ± 1425.3	-48.9 ± 27.9
PHE	CA	43.7 ± 13.1	136.4 ± 41.5	236.4 ± 72.4	-417.5 ± 127.4	-4.3 ± 2.5
PHE	CB	33.9 ± 10.2	108.4 ± 32.1	186.2 ± 56.0	-330.0 ± 98.5	-3.4 ± 1.9
PHE	CD1	5.9 ± 2.0	17.1 ± 6.2	29.9 ± 11.1	-52.6 ± 19.3	-0.1 ± 0.4
PHE	CD2	5.9 ± 2.1	17.1 ± 6.4	29.9 ± 11.4	-52.6 ± 19.9	-0.1 ± 0.4
PHE	CE1	-25.7 ± 7.7	-80.5 ± 24.1	-141.0 ± 42.1	247.1 ± 74.1	2.4 ± 1.5
PHE	CE2	-25.7 ± 7.6	-80.5 ± 24.0	-140.9 ± 41.9	247.0 ± 73.7	2.4 ± 1.5
PHE	CG	21.5 ± 6.4	68.1 ± 20.3	117.8 ± 35.5	-208.3 ± 62.4	-2.2 ± 1.2
PHE	CZ	16.7 ± 5.4	54.1 ± 16.9	91.5 ± 29.8	-160.8 ± 52.2	-1.7 ± 1.0
PHE	N	-132.0 ± 43.7	-435.2 ± 135.8	-749.0 ± 236.2	1261.0 ± 415.3	13.9 ± 8.0
PHE	O	-430.8 ± 130.1	-1348.0 ± 412.0	-2351.0 ± 716.8	4097.0 ± 1261.4	42.6 ± 24.6
PRO	C	466.8 ± 146.1	1462.0 ± 462.7	2495.0 ± 799.8	-4508.0 ± 1409.1	-47.5 ± 27.6
PRO	CA	-50.6 ± 18.4	-155.9 ± 57.7	-275.6 ± 100.5	480.2 ± 176.8	4.9 ± 3.2
PRO	CB	47.1 ± 15.2	147.5 ± 47.9	255.3 ± 83.2	-453.4 ± 146.4	-4.9 ± 2.8
PRO	CD	81.6 ± 25.7	253.7 ± 81.4	439.9 ± 141.3	-783.5 ± 248.7	-8.3 ± 4.8
PRO	CG	51.5 ± 16.7	161.5 ± 52.8	283.0 ± 91.7	-499.8 ± 161.4	-5.3 ± 3.1
PRO	N	-188.8 ± 61.3	-602.6 ± 194.1	-1033.0 ± 335.7	1808.0 ± 590.9	19.5 ± 11.4
PRO	O	-422.8 ± 128.2	-1279.0 ± 407.3	-2253.0 ± 702.3	3927.0 ± 1236.6	41.4 ± 24.2
SER	C	464.2 ± 146.2	1462.0 ± 462.7	2506.0 ± 804.6	-4501.0 ± 1416.7	-41.9 ± 28.4
SER	CA	61.0 ± 19.3	190.0 ± 61.0	330.5 ± 106.2	-582.9 ± 186.9	-5.4 ± 3.7
SER	CB	141.3 ± 45.7	447.5 ± 144.4	770.5 ± 251.8	-1368.0 ± 443.0	-12.9 ± 8.8
SER	N	-125.3 ± 46.8	-417.3 ± 146.8	-714.1 ± 254.9	1178.0 ± 445.3	11.8 ± 8.3
SER	O	-414.5 ± 127.8	-1295.0 ± 406.3	-2256.0 ± 705.3	3946.0 ± 1240.9	36.5 ± 24.9
SER	OG	-131.9 ± 41.7	-427.1 ± 132.1	-710.2 ± 230.0	1285.0 ± 404.6	12.2 ± 8.3
THR	C	478.3 ± 147.0	1472.0 ± 463.5	2531.0 ± 803.5	-4563.0 ± 1414.0	-44.3 ± 29.2
THR	CA	60.9 ± 19.1	188.1 ± 60.1	328.0 ± 104.4	-578.8 ± 183.6	-5.6 ± 3.7
THR	CB	217.2 ± 69.2	677.8 ± 217.9	1170.0 ± 378.9	-2076.0 ± 666.3	-20.8 ± 13.7
THR	CG2	-43.3 ± 14.0	-138.4 ± 43.9	-230.1 ± 76.5	424.3 ± 134.4	4.2 ± 2.7
THR	N	-141.2 ± 45.4	-421.6 ± 140.7	-723.7 ± 243.7	1195.0 ± 427.0	12.5 ± 8.4
THR	O	-405.6 ± 128.5	-1304.0 ± 406.4	-2276.0 ± 703.5	3998.0 ± 1237.2	38.5 ± 25.5
THR	OG1	-440.8 ± 141.8	-1405.0 ± 447.5	-2422.0 ± 777.4	4293.0 ± 1366.9	42.2 ± 27.9
TRP	C	486.7 ± 147.1	1543.0 ± 473.3	2657.0 ± 821.9	-4795.0 ± 1450.3	-46.0 ± 29.1
TRP	CA	44.5 ± 13.3	140.3 ± 42.8	243.6 ± 74.5	-432.4 ± 131.4	-4.2 ± 2.6
TRP	CB	32.8 ± 9.8	105.2 ± 31.6	183.2 ± 55.0	-325.3 ± 97.0	-3.1 ± 2.0
TRP	CD1	45.9 ± 13.9	140.1 ± 44.6	251.9 ± 78.0	-443.2 ± 137.5	-4.0 ± 2.7
TRP	CD2	56.6 ± 16.9	177.4 ± 54.3	314.6 ± 94.5	-552.7 ± 166.7	-5.3 ± 3.3
TRP	CE2	121.9 ± 36.5	376.8 ± 117.6	671.1 ± 204.8	-1180.0 ± 361.3	-11.1 ± 7.1
TRP	CE3	-278.0 ± 82.6	-877.4 ± 266.2	-1531.0 ± 462.7	2695.0 ± 816.5	25.1 ± 16.2
TRP	CG	-72.2 ± 21.6	-230.3 ± 69.6	-399.2 ± 121.2	707.3 ± 213.9	6.9 ± 4.3
TRP	CH2	-2.9 ± 1.3	-5.4 ± 3.8	-16.9 ± 7.0	27.3 ± 12.1	0.2 ± 0.3
TRP	CZ2	-66.5 ± 19.8	-224.9 ± 63.8	-371.0 ± 111.0	648.4 ± 195.9	6.1 ± 3.9
TRP	CZ3	-30.0 ± 9.0	-98.5 ± 28.9	-166.7 ± 50.3	292.6 ± 88.8	2.7 ± 1.8
TRP	N	-132.2 ± 44.2	-443.3 ± 139.7	-763.5 ± 242.8	1285.0 ± 427.3	13.0 ± 8.3
TRP	NE1	-12.9 ± 4.3	-38.0 ± 13.7	-66.7 ± 24.2	114.1 ± 42.7	1.0 ± 0.9
TRP	O	-434.3 ± 129.3	-1369.0 ± 417.4	-2385.0 ± 723.7	4173.0 ± 1276.5	40.1 ± 25.6

Table 9: Continued

Residue		Aliphatic	Aromatic	Polar Carbon	Polar	Alpha Carbon
TYR	C	485.7 ± 144.6	1533.0 ± 458.7	2629.0 ± 797.8	-4744.0 ± 1406.0	-47.1 ± 29.1
TYR	CA	39.7 ± 11.7	125.4 ± 37.3	215.1 ± 64.9	-383.6 ± 114.4	-3.9 ± 2.4
TYR	CB	48.3 ± 14.3	155.8 ± 45.3	265.1 ± 79.0	-470.6 ± 139.1	-4.7 ± 2.9
TYR	CD1	-18.5 ± 5.9	-61.4 ± 18.7	-104.7 ± 32.7	178.4 ± 57.6	1.8 ± 1.1
TYR	CD2	-18.5 ± 5.9	-61.4 ± 18.8	-104.8 ± 32.8	178.5 ± 57.7	1.8 ± 1.1
TYR	CE1	-35.3 ± 10.5	-112.0 ± 33.2	-194.4 ± 57.9	346.0 ± 102.0	3.4 ± 2.1
TYR	CE2	-35.3 ± 10.5	-111.9 ± 33.3	-194.3 ± 57.9	348.5 ± 102.1	3.4 ± 2.1
TYR	CG	-1.2 ± 0.4	-3.9 ± 1.2	-6.8 ± 2.1	12.0 ± 3.6	0.1 ± 0.1
TYR	CZ	182.3 ± 55.4	569.7 ± 176.1	994.7 ± 307.0	-1751.0 ± 540.9	-17.6 ± 10.8
TYR	N	-132.7 ± 42.9	-439.5 ± 134.0	-751.4 ± 232.8	1268.0 ± 409.5	13.3 ± 8.3
TYR	O	-433.0 ± 127.8	-1362.0 ± 407.7	-2368.0 ± 706.6	4134.0 ± 1244.7	41.0 ± 25.6
TYR	OH	-84.8 ± 26.3	-264.1 ± 83.4	-462.0 ± 145.4	812.7 ± 255.9	8.2 ± 5.1
VAL	C	516.5 ± 146.6	1520.0 ± 460.3	2621.0 ± 801.2	-4716.0 ± 1408.7	-49.5 ± 28.7
VAL	CA	-8.1 ± 3.0	-29.6 ± 9.5	-54.5 ± 16.5	92.7 ± 28.9	0.9 ± 0.7
VAL	CB	258.7 ± 77.0	804.0 ± 241.5	1396.0 ± 421.6	-2469.0 ± 741.0	-26.5 ± 15.3
VAL	CG1	-79.3 ± 25.7	-271.4 ± 80.6	-455.3 ± 140.7	824.5 ± 247.3	8.8 ± 5.1
VAL	CG2	-79.3 ± 25.7	-271.7 ± 80.4	-458.7 ± 140.6	822.0 ± 246.8	8.8 ± 5.1
VAL	N	-164.8 ± 44.0	-436.8 ± 135.1	-751.4 ± 234.7	1263.0 ± 411.8	14.0 ± 8.2
VAL	O	-416.4 ± 128.5	-1349.0 ± 405.5	-2357.0 ± 704.6	4101.0 ± 1238.5	43.0 ± 25.2

All units in kcal/mol.

Lennard-Jones Attractive Interactions

Figure 7 shows the total Lennard-Jones attractive interactions normalized to the protein size (number of heavy atoms). The strongest interaction is the aromatic-aromatic interactions, representing the stacking of aromatic rings. Because the distance between any two alpha carbons is limited by the structure of the backbone, the interactions between the alpha carbons is relatively weak. Interactions between alpha carbons and aromatic carbons are also weak, indicating that aromatic groups do not interact significantly with the backbone.

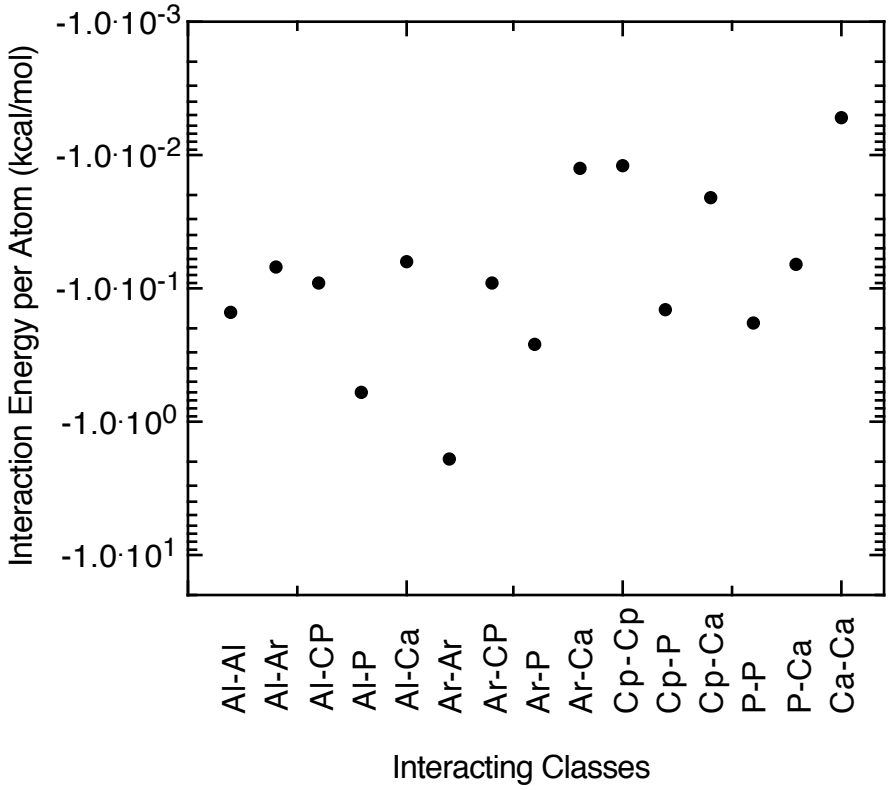


Figure 7: Total Lennard-Jones attractive interactions divided into types. The interaction energy has been normalized for the size (number of atoms) of the protein.

Figure 8 shows the average Lennard-Jones attractive interactions for each amino acid. There appears to be a strong interaction between the cysteine and the aliphatic groups. However, this is an artifact of not accounting for disulphide bond formation during the calculation. Figure 5 also shows strong interactions between polar and aliphatic groups with all the amino acids, but tryptophan and tyrosine are the most significant.

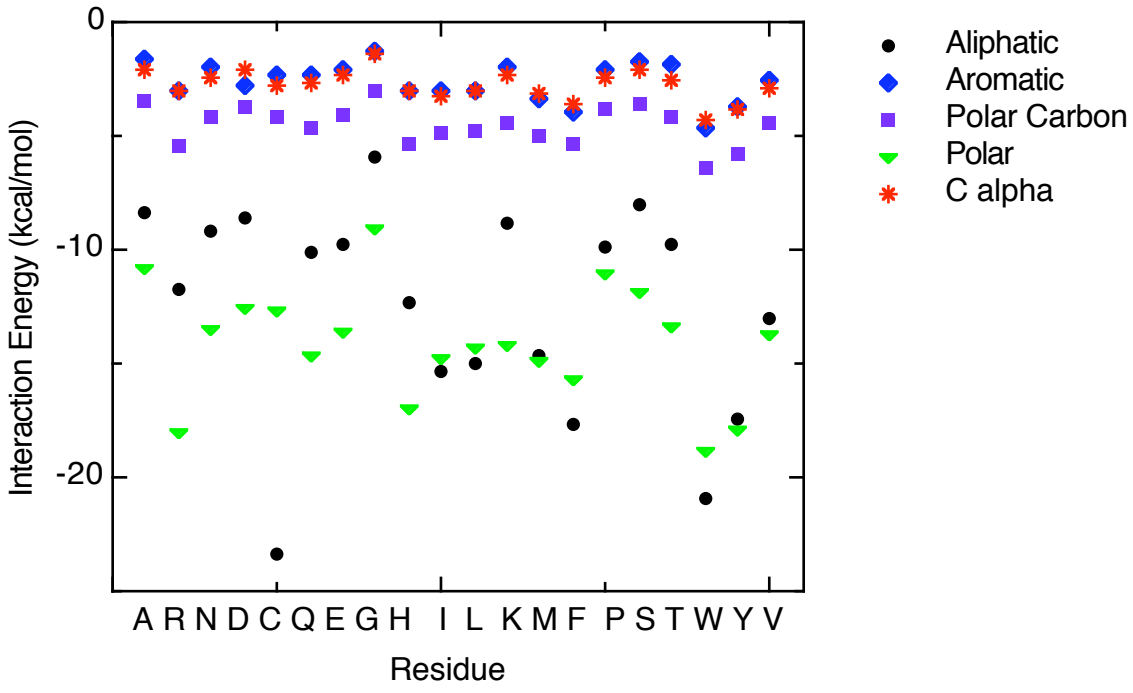


Figure 8: Average Lennard-Jones attractive interaction between each residue and other atom groups in the protein.

Figure 9 shows the average contribution of individual atom groups to the overall Lennard-Jones attractive interactions of the protein. Table 10 shows the data used to make the Figure 9. Since the interaction depends on the well depth, the groups in each panel have been sorted by the well depth. In each classification, a few interesting atom groups appear significantly different.

In panel A, there appears to be a significant interaction between aliphatic groups and the cysteine Sγ. As mentioned above, this is an artifact of not accounting for disulphide bond formation during the calculation. In addition, lysine Cε has a significant interaction with polar groups. In other charged side chains, the carbon attached to the polar atoms is classified as a polar carbon due to the existence of resonance structures. This is not the case for lysine. Polar interactions bring the lysine Cε into close proximity with polar groups, increasing the favorable van der Waals interaction. In panel B, several groups form strong interactions with other aromatic groups. All of these belong to tryptophan and indicate that aromatic stacking interactions involving tryptophan are more significant than those involving tyrosine or phenylalanine. In panel E, lysine Nζ stands out as having strong interactions with aromatic, polar and aliphatic groups. There is a significant deviation in these values for lysine Nζ and can be accounted for by errors associated with the position of the Nζ and the associated hydrogen atoms in a subset of the crystal structures.

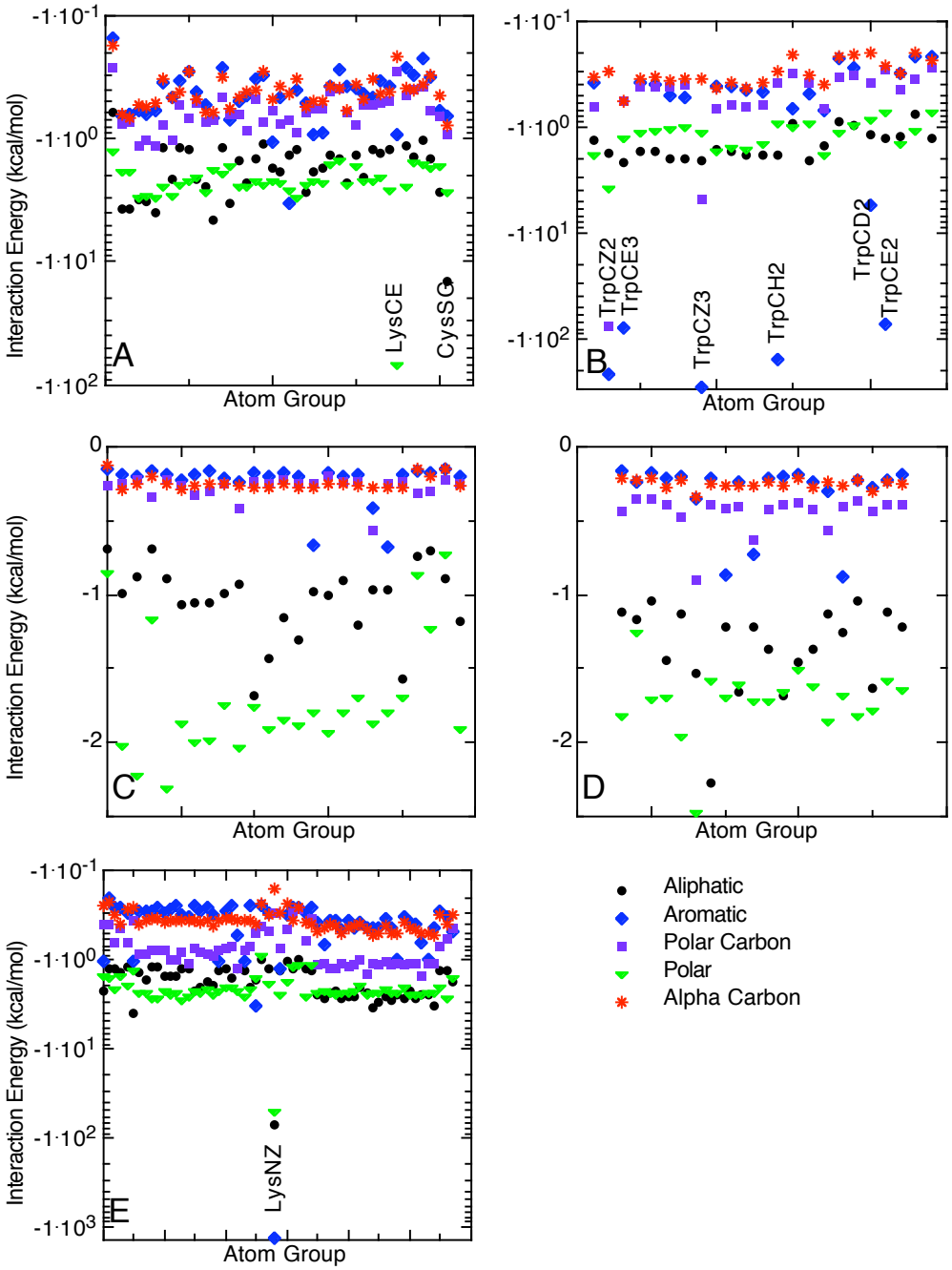


Figure 9: Lennard-Jones attractive interactions between individual groups and all other groups. A) Aliphatic Groups B) Aromatic Groups C) Polar Carbons D) Alpha Carbon E) Polar Groups. Groups in each panel have been sorted by the van der Waals well depth parameter of AMBER 99.

Table 10: Lennard-Jones attractive interactions between the given atom group and class

Residue		Aliphatic	Aromatic	Polar Carbon	Polar	Alpha Carbon
ALA	C	-0.99 ± 0.42	-0.19 ± 0.21	-0.25 ± 0.11	-2.03 ± 1.36	-0.28 ± 0.09
ALA	CA	-1.38 ± 0.70	-0.24 ± 0.36	-0.43 ± 0.17	-1.62 ± 0.57	-0.27 ± 0.12
ALA	CB	-2.42 ± 1.61	-0.53 ± 0.77	-0.73 ± 0.39	-2.75 ± 1.05	-0.61 ± 0.28
ALA	N	-1.37 ± 0.58	-0.28 ± 0.29	-0.88 ± 1.00	-2.34 ± 0.93	-0.40 ± 0.13
ALA	O	-2.44 ± 0.97	-0.39 ± 0.55	-1.15 ± 0.44	-2.37 ± 1.29	-0.48 ± 0.26
ARG	C	-1.19 ± 0.46	-0.19 ± 0.22	-0.25 ± 0.11	-1.91 ± 0.62	-0.26 ± 0.08
ARG	CA	-1.37 ± 0.59	-0.22 ± 0.31	-0.43 ± 0.16	-1.72 ± 0.57	-0.23 ± 0.10
ARG	CB	-1.34 ± 0.79	-3.34 ± 288.20	-0.69 ± 0.30	-2.64 ± 1.02	-0.43 ± 0.16
ARG	CD	-1.24 ± 0.91	-0.45 ± 0.59	-0.53 ± 0.39	-2.21 ± 1.16	-0.33 ± 0.22
ARG	CG	-1.23 ± 0.88	-0.39 ± 0.69	-0.90 ± 0.36	-2.98 ± 1.08	-0.32 ± 0.17
ARG	CZ	-0.89 ± 0.47	-0.15 ± 0.28	-0.23 ± 0.16	-0.73 ± 0.48	-0.15 ± 0.11
ARG	N	-1.71 ± 0.61	-0.28 ± 0.30	-0.85 ± 0.38	-2.33 ± 0.93	-0.37 ± 0.12
ARG	NE	-4.02 ± 288.69	-1.04 ± 82.02	-0.37 ± 0.26	-1.33 ± 1.07	-0.27 ± 0.16
ARG	NH1	-2.22 ± 1.23	-1.06 ± 82.02	-0.40 ± 0.35	-1.55 ± 1.48	-0.24 ± 0.22
ARG	NH2	-1.28 ± 0.92	-0.21 ± 0.42	-0.41 ± 0.36	-1.65 ± 1.51	-0.22 ± 0.21
ARG	O	-2.72 ± 1.03	-0.67 ± 27.17	-1.12 ± 0.47	-2.29 ± 1.48	-0.45 ± 0.27
ASN	C	-0.88 ± 0.38	-0.19 ± 0.22	-0.23 ± 0.11	-2.23 ± 0.65	-0.25 ± 0.08
ASN	CA	-1.04 ± 0.54	-0.23 ± 0.29	-0.37 ± 0.16	-1.83 ± 0.57	-0.22 ± 0.11
ASN	CB	-1.16 ± 0.86	-0.34 ± 0.55	-0.53 ± 0.30	-2.38 ± 1.02	-0.41 ± 0.18
ASN	CG	-0.69 ± 0.47	-0.17 ± 0.24	-0.33 ± 0.16	-1.17 ± 0.51	-0.20 ± 0.11
ASN	N	-1.21 ± 0.54	-0.28 ± 0.29	-0.79 ± 0.38	-2.80 ± 1.06	-0.35 ± 0.11
ASN	NE2	-1.28 ± 1.02	-0.26 ± 0.51	-0.64 ± 0.46	-2.11 ± 1.52	-0.32 ± 0.25
ASN	O	-2.32 ± 0.94	-0.36 ± 0.42	-1.27 ± 0.52	-2.38 ± 1.19	-0.41 ± 0.27
ASN	OE1	-1.34 ± 0.95	-0.29 ± 0.40	-0.71 ± 0.50	-2.03 ± 1.31	-0.31 ± 0.28
ASP	C	-0.89 ± 0.38	-0.18 ± 0.21	-0.22 ± 0.11	-2.31 ± 0.65	-0.25 ± 0.08
ASP	CA	-1.05 ± 0.54	-0.18 ± 0.25	-0.35 ± 0.15	-1.72 ± 0.55	-0.21 ± 0.10
ASP	CB	-0.62 ± 0.54	-0.15 ± 0.27	-0.27 ± 0.16	-1.27 ± 0.55	-0.18 ± 0.09
ASP	CG	-0.70 ± 0.48	-0.18 ± 0.22	-0.30 ± 0.16	-1.23 ± 0.58	-0.20 ± 0.12
ASP	N	-1.24 ± 0.55	-0.27 ± 0.30	-0.77 ± 1.15	-2.68 ± 1.07	-0.34 ± 0.11
ASP	O	-2.31 ± 0.99	-0.36 ± 0.40	-1.30 ± 0.52	-2.46 ± 1.20	-0.42 ± 0.27
ASP	OD1	-1.33 ± 0.95	-1.27 ± 102.78	-0.76 ± 0.53	-2.47 ± 2.25	-0.30 ± 0.27
ASP	OD2	-1.05 ± 0.88	-0.31 ± 1.06	-0.47 ± 0.62	-1.75 ± 1.39	-0.24 ± 0.26
CYS	C	-1.07 ± 0.42	-0.22 ± 0.24	-0.26 ± 0.12	-1.87 ± 2.36	-0.28 ± 0.10
CYS	CA	-1.63 ± 0.69	-0.27 ± 0.33	-0.44 ± 0.15	-1.79 ± 0.60	-0.30 ± 0.12
CYS	CB	-2.68 ± 1.92	-0.50 ± 0.66	-0.61 ± 0.30	-2.34 ± 0.83	-0.54 ± 0.19
CYS	N	-1.51 ± 0.60	-0.30 ± 0.31	-0.78 ± 0.36	-2.21 ± 0.88	-0.38 ± 0.11
CYS	O	-2.74 ± 1.00	-0.46 ± 0.43	-1.12 ± 0.45	-2.33 ± 1.19	-0.51 ± 0.28
CYS	SG	-14.22 ± 14.46	-0.65 ± 0.73	-0.91 ± 0.41	-2.69 ± 1.06	-0.78 ± 0.37
GLN	C	-1.05 ± 0.41	-0.18 ± 0.22	-0.32 ± 0.15	-2.00 ± 1.59	-0.26 ± 0.09
GLN	CA	-1.13 ± 0.52	-0.20 ± 0.29	-0.48 ± 0.21	-1.96 ± 0.65	-0.23 ± 0.10
GLN	CB	-1.21 ± 0.89	-0.38 ± 0.57	-0.49 ± 0.25	-2.63 ± 1.26	-0.42 ± 0.16
GLN	CD	-0.74 ± 0.52	-0.16 ± 0.26	-0.31 ± 0.17	-0.87 ± 0.45	-0.15 ± 0.11
GLN	CG	-1.19 ± 0.92	-0.35 ± 0.58	-0.77 ± 0.37	-2.46 ± 1.17	-0.32 ± 0.18
GLN	N	-1.53 ± 0.61	-0.27 ± 0.27	-1.02 ± 1.52	-2.46 ± 1.04	-0.37 ± 0.13
GLN	NE2	-1.40 ± 1.42	-0.25 ± 0.49	-0.43 ± 0.36	-1.56 ± 1.39	-0.40 ± 0.29
GLN	O	-2.60 ± 1.04	-0.37 ± 0.39	-1.19 ± 0.52	-2.35 ± 1.43	-0.44 ± 0.27
GLN	OE1	-1.33 ± 0.95	-0.26 ± 0.46	-0.34 ± 0.29	-1.15 ± 1.04	-0.40 ± 0.31
GLU	C	-1.05 ± 0.41	-0.17 ± 0.19	-0.30 ± 0.13	-1.99 ± 0.62	-0.25 ± 0.08
GLU	CA	-1.12 ± 0.53	-0.17 ± 0.26	-0.43 ± 0.19	-1.83 ± 0.62	-0.21 ± 0.09
GLU	CB	-1.13 ± 0.86	-0.27 ± 0.44	-0.44 ± 0.25	-2.45 ± 0.83	-0.39 ± 0.16
GLU	CD	-0.69 ± 0.49	-0.15 ± 0.23	-0.27 ± 0.16	-0.85 ± 0.51	-0.12 ± 0.10
GLU	CG	-1.22 ± 1.09	-0.29 ± 0.56	-0.68 ± 0.34	-2.19 ± 1.10	-0.28 ± 0.17
GLU	N	-1.55 ± 1.22	-0.25 ± 0.26	-0.99 ± 0.44	-2.36 ± 1.00	-0.36 ± 0.11
GLU	O	-2.60 ± 1.07	-0.44 ± 11.89	-1.15 ± 0.48	-2.27 ± 1.26	-0.42 ± 0.26
GLU	OE1	-1.29 ± 0.87	-0.25 ± 0.40	-0.30 ± 0.27	-1.21 ± 1.18	-0.37 ± 0.30
GLU	OE2	-1.02 ± 0.97	-0.26 ± 0.45	-0.27 ± 0.26	-1.16 ± 1.18	-0.27 ± 0.24

Table 10: Continued

Residue		Aliphatic	Aromatic	Polar Carbon	Polar	Alpha Carbon
GLY	C	-1.00 ± 0.49	-0.22 ± 0.29	-0.25 ± 0.14	-1.75 ± 1.77	-0.26 ± 0.11
GLY	CA	-1.53 ± 1.00	-0.35 ± 0.59	-0.90 ± 0.32	-2.48 ± 1.02	-0.34 ± 0.20
GLY	N	-1.28 ± 0.69	-0.31 ± 0.39	-0.82 ± 1.30	-2.92 ± 90.72	-0.37 ± 0.14
GLY	O	-2.07 ± 1.09	-0.38 ± 0.45	-1.02 ± 0.47	-1.99 ± 1.11	-0.40 ± 0.28
HIS	C	-0.92 ± 0.38	-0.23 ± 0.23	-0.42 ± 0.19	-2.04 ± 0.67	-0.26 ± 0.08
HIS	CA	-1.13 ± 0.54	-0.30 ± 0.31	-0.57 ± 0.18	-1.87 ± 0.90	-0.24 ± 0.10
HIS	CB	-1.50 ± 0.99	-0.49 ± 0.63	-0.62 ± 0.32	-2.45 ± 0.93	-0.46 ± 0.18
HIS	CD2	-1.32 ± 0.86	-0.39 ± 1.28	-0.65 ± 0.39	-1.87 ± 1.26	-0.33 ± 0.20
HIS	CE1	-1.25 ± 0.88	-0.31 ± 0.62	-0.44 ± 0.36	-1.42 ± 2.59	-0.31 ± 0.19
HIS	CG	-0.76 ± 0.46	-0.22 ± 0.28	-0.36 ± 0.17	-1.11 ± 0.39	-0.20 ± 0.09
HIS	N	-1.26 ± 0.57	-0.34 ± 0.32	-1.12 ± 0.42	-2.57 ± 0.94	-0.36 ± 0.11
HIS	ND1	-1.23 ± 0.76	-0.31 ± 0.43	-0.64 ± 0.35	-1.96 ± 1.11	-0.27 ± 0.18
HIS	NE2	-1.27 ± 0.86	-0.31 ± 0.52	-0.48 ± 0.36	-1.88 ± 20.56	-0.32 ± 0.19
HIS	O	-2.40 ± 0.95	-0.45 ± 0.43	-1.44 ± 0.57	-2.44 ± 1.15	-0.47 ± 0.28
ILE	C	-1.68 ± 0.38	-0.18 ± 0.19	-0.24 ± 0.10	-1.76 ± 0.61	-0.28 ± 0.07
ILE	CA	-1.69 ± 0.58	-0.20 ± 0.24	-0.39 ± 0.12	-1.67 ± 0.51	-0.26 ± 0.09
ILE	CB	-1.43 ± 0.93	-0.28 ± 0.35	-0.39 ± 0.15	-1.52 ± 0.45	-0.39 ± 0.11
ILE	CD1	-4.56 ± 1.80	-0.68 ± 0.94	-0.70 ± 0.41	-1.82 ± 1.03	-0.60 ± 0.32
ILE	CG1	-2.29 ± 1.01	-0.45 ± 0.57	-0.85 ± 0.32	-2.48 ± 0.84	-0.42 ± 0.17
ILE	CG2	-3.97 ± 1.43	-0.59 ± 0.77	-1.13 ± 0.36	-3.02 ± 0.91	-0.52 ± 0.24
ILE	N	-2.30 ± 0.66	-0.25 ± 0.26	-0.75 ± 0.34	-2.23 ± 0.82	-0.37 ± 0.10
ILE	O	-3.43 ± 0.94	-0.44 ± 0.39	-1.11 ± 0.43	-2.38 ± 1.04	-0.53 ± 0.26
LEU	C	-1.43 ± 0.52	-0.19 ± 0.21	-0.25 ± 0.10	-1.91 ± 0.61	-0.28 ± 0.08
LEU	CA	-2.28 ± 0.57	-0.22 ± 0.27	-0.39 ± 0.12	-1.58 ± 0.45	-0.25 ± 0.09
LEU	CB	-2.09 ± 0.85	-0.41 ± 0.53	-0.51 ± 0.22	-2.08 ± 0.61	-0.47 ± 0.14
LEU	CD1	-3.68 ± 1.60	-0.65 ± 0.85	-0.74 ± 0.44	-1.83 ± 1.02	-0.62 ± 0.31
LEU	CD2	-3.67 ± 1.63	-0.63 ± 0.86	-0.72 ± 0.42	-1.83 ± 1.07	-0.68 ± 0.31
LEU	CG	-1.46 ± 0.67	-0.32 ± 0.38	-0.59 ± 0.21	-1.75 ± 0.54	-0.30 ± 0.11
LEU	N	-2.09 ± 0.64	-0.28 ± 0.30	-0.84 ± 1.04	-2.38 ± 0.86	-0.39 ± 0.11
LEU	O	-3.01 ± 1.07	-0.43 ± 0.41	-1.15 ± 0.45	-2.42 ± 1.10	-0.50 ± 0.27
LYS	C	-1.16 ± 0.46	-0.17 ± 0.21	-0.23 ± 0.11	-1.85 ± 1.26	-0.25 ± 0.08
LYS	CA	-1.45 ± 1.39	-0.19 ± 0.30	-0.37 ± 0.15	-1.50 ± 0.51	-0.22 ± 0.10
LYS	CB	-1.47 ± 0.85	-0.32 ± 1.22	-0.48 ± 0.24	-2.17 ± 1.15	-0.40 ± 0.15
LYS	CD	-1.18 ± 1.42	-0.26 ± 0.49	-0.47 ± 0.34	-1.95 ± 1.50	-0.31 ± 0.23
LYS	CE	-0.93 ± 0.77	-0.92 ± 0.43	-0.28 ± 0.24	-69.44 ± 6.81	-0.22 ± 0.17
LYS	CG	-1.11 ± 0.83	-0.30 ± 0.52	-0.72 ± 0.33	-2.38 ± 0.95	-0.29 ± 0.16
LYS	N	-1.74 ± 0.71	-0.26 ± 0.28	-0.85 ± 1.64	-2.20 ± 0.95	-0.36 ± 0.13
LYS	NZ	-70.73 ± 7.10	-1341.00 ± 1.4E+05	-0.30 ± 0.34	-52.65 ± 10.87	-0.16 ± 0.17
LYS	O	-2.64 ± 1.68	-0.34 ± 0.38	-1.07 ± 0.48	-2.12 ± 1.25	-0.41 ± 0.27
MET	C	-1.30 ± 0.46	-0.20 ± 0.22	-0.25 ± 0.11	-1.88 ± 0.65	-0.28 ± 0.09
MET	CA	-1.66 ± 0.65	-0.24 ± 0.31	-0.41 ± 0.15	-1.61 ± 0.58	-0.26 ± 0.10
MET	CB	-2.07 ± 0.98	-0.43 ± 0.58	-0.54 ± 0.27	-2.18 ± 0.76	-0.47 ± 0.17
MET	CE	-3.30 ± 2.06	-0.71 ± 1.11	-0.63 ± 0.47	-1.64 ± 1.09	-0.57 ± 0.34
MET	CG	-1.82 ± 1.06	-0.46 ± 0.63	-0.74 ± 0.33	-2.29 ± 0.87	-0.37 ± 0.18
MET	N	-1.98 ± 0.70	-0.31 ± 0.33	-0.89 ± 0.41	-2.50 ± 1.01	-0.41 ± 0.13
MET	O	-2.93 ± 1.06	-0.45 ± 0.42	-1.15 ± 0.46	-2.38 ± 1.15	-0.50 ± 0.27
MET	SD	-2.76 ± 1.39	-0.59 ± 0.70	-0.66 ± 0.38	-1.64 ± 0.84	-0.44 ± 0.27

Table 10: Continued

Residue		Aliphatic	Aromatic	Polar Carbon	Polar	Alpha Carbon
PHE	C	-0.98 ± 0.38	-0.67 ± 0.41	-0.25 ± 0.11	-1.80 ± 0.62	-0.27 ± 0.08
PHE	CA	-1.25 ± 0.51	-0.89 ± 0.40	-0.41 ± 0.15	-1.68 ± 0.56	-0.26 ± 0.10
PHE	CB	-1.82 ± 0.91	-0.91 ± 0.65	-0.56 ± 0.28	-2.23 ± 0.74	-0.49 ± 0.17
PHE	CD1	-1.81 ± 0.78	-0.45 ± 0.50	-0.64 ± 0.31	-1.62 ± 0.75	-0.43 ± 0.18
PHE	CD2	-1.81 ± 0.79	-0.46 ± 0.51	-0.61 ± 0.30	-1.46 ± 0.64	-0.38 ± 0.17
PHE	CE1	-2.00 ± 0.99	-0.50 ± 0.55	-0.41 ± 0.26	-1.03 ± 0.60	-0.37 ± 0.19
PHE	CE2	-2.02 ± 0.99	-0.52 ± 0.57	-0.40 ± 0.26	-0.99 ± 0.57	-0.36 ± 0.18
PHE	CG	-0.97 ± 0.41	-0.27 ± 0.27	-0.32 ± 0.14	-0.98 ± 0.30	-0.21 ± 0.08
PHE	CZ	-2.10 ± 1.01	-0.49 ± 0.57	-0.38 ± 0.28	-0.93 ± 0.59	-0.32 ± 0.19
PHE	N	-1.32 ± 0.54	-1.04 ± 0.53	-0.78 ± 0.36	-2.21 ± 0.87	-0.37 ± 0.11
PHE	O	-2.50 ± 0.92	-1.01 ± 0.64	-1.11 ± 0.44	-2.35 ± 1.08	-0.51 ± 0.27
PRO	C	-1.01 ± 0.40	-0.18 ± 0.20	-0.20 ± 0.11	-1.93 ± 0.51	-0.25 ± 0.11
PRO	CA	-1.17 ± 0.69	-0.24 ± 0.39	-0.35 ± 0.20	-1.26 ± 0.58	-0.23 ± 0.13
PRO	CB	-1.35 ± 1.07	-0.37 ± 0.61	-0.41 ± 0.32	-1.59 ± 0.78	-0.38 ± 0.19
PRO	CD	-2.29 ± 1.06	-0.38 ± 0.64	-0.62 ± 0.44	-2.35 ± 1.21	-0.58 ± 0.20
PRO	CG	-1.36 ± 1.15	-0.39 ± 0.68	-0.77 ± 0.39	-1.67 ± 1.05	-0.37 ± 0.21
PRO	N	-0.99 ± 0.44	-0.25 ± 0.29	-0.42 ± 0.21	-0.90 ± 0.43	-0.23 ± 0.08
PRO	O	-2.73 ± 1.06	-0.34 ± 0.42	-1.15 ± 0.49	-2.04 ± 1.07	-0.38 ± 0.24
SER	C	-0.90 ± 0.40	-0.20 ± 0.22	-0.23 ± 0.12	-1.80 ± 0.59	-0.25 ± 0.09
SER	CA	-1.12 ± 0.60	-0.22 ± 0.33	-0.39 ± 0.18	-1.59 ± 0.63	-0.24 ± 0.12
SER	CB	-1.33 ± 1.03	-0.33 ± 0.54	-0.52 ± 0.32	-2.08 ± 0.97	-0.43 ± 0.21
SER	N	-1.24 ± 0.57	-0.28 ± 0.38	-0.76 ± 0.38	-2.07 ± 0.95	-0.34 ± 0.12
SER	O	-2.21 ± 0.97	-0.39 ± 0.41	-1.09 ± 0.50	-2.50 ± 1.18	-0.44 ± 0.28
SER	OG	-1.34 ± 1.17	-0.34 ± 0.48	-0.60 ± 0.35	-2.74 ± 1.50	-0.40 ± 0.27
THR	C	-1.21 ± 0.45	-0.19 ± 0.22	-0.23 ± 0.11	-1.70 ± 0.59	-0.26 ± 0.08
THR	CA	-1.22 ± 0.54	-0.19 ± 0.25	-0.38 ± 0.15	-1.65 ± 0.60	-0.25 ± 0.11
THR	CB	-1.03 ± 0.66	-0.22 ± 0.32	-0.37 ± 0.20	-1.59 ± 0.62	-0.36 ± 0.14
THR	CG2	-2.11 ± 1.54	-0.45 ± 1.51	-1.03 ± 0.43	-2.93 ± 1.24	-0.46 ± 0.27
THR	N	-1.61 ± 0.60	-0.25 ± 0.27	-0.72 ± 0.36	-2.06 ± 0.90	-0.34 ± 0.11
THR	O	-2.70 ± 1.04	-0.40 ± 0.40	-1.10 ± 0.49	-2.50 ± 2.37	-0.46 ± 0.28
THR	OG1	-1.30 ± 0.95	-0.30 ± 0.41	-0.61 ± 0.31	-2.56 ± 2.51	-0.39 ± 0.25
TRP	C	-0.97 ± 0.38	-0.41 ± 0.28	-0.57 ± 0.25	-1.88 ± 0.64	-0.27 ± 0.08
TRP	CA	-1.22 ± 0.51	-0.73 ± 0.38	-0.63 ± 0.16	-1.72 ± 0.52	-0.26 ± 0.11
TRP	CB	-1.72 ± 0.92	-1.04 ± 0.62	-0.59 ± 0.26	-2.21 ± 0.74	-0.48 ± 0.17
TRP	CD1	-1.51 ± 0.75	-0.70 ± 0.47	-0.67 ± 0.32	-1.84 ± 0.81	-0.39 ± 0.19
TRP	CD2	-1.16 ± 0.50	-5.54 ± 0.33	-0.38 ± 0.17	-0.86 ± 0.30	-0.20 ± 0.10
TRP	CE2	-1.27 ± 0.59	-73.75 ± 3.21	-0.29 ± 0.15	-0.72 ± 0.32	-0.26 ± 0.12
TRP	CE3	-2.18 ± 0.82	-77.54 ± 4.25	-0.57 ± 0.32	-1.29 ± 0.65	-0.56 ± 0.23
TRP	CG	-0.92 ± 0.42	-0.67 ± 0.27	-0.32 ± 0.13	-1.02 ± 0.31	-0.21 ± 0.09
TRP	CH2	-1.88 ± 1.01	-153.80 ± 5.07	-0.39 ± 0.26	-0.92 ± 0.58	-0.30 ± 0.20
TRP	CZ2	-1.77 ± 0.96	-216.10 ± 31.76	-75.62 ± 3.27	-3.86 ± 0.65	-0.29 ± 0.20
TRP	CZ3	-2.04 ± 0.98	-280.70 ± 32.90	-4.76 ± 0.32	-1.12 ± 0.60	-0.35 ± 0.21
TRP	N	-1.30 ± 0.54	-0.53 ± 0.36	-1.27 ± 0.43	-2.23 ± 0.87	-0.37 ± 0.11
TRP	NE1	-1.67 ± 0.84	-3.38 ± 0.53	-0.50 ± 0.30	-1.64 ± 1.01	-0.39 ± 0.21
TRP	O	-2.47 ± 0.94	-0.66 ± 0.47	-1.55 ± 0.63	-2.44 ± 1.15	-0.50 ± 0.27

Table 10: Continued

Residue	Aliphatic	Aromatic	Polar Carbon	Polar	Alpha Carbon
TYR C	-0.97 ± 0.38	-0.68 ± 0.41	-0.25 ± 0.11	-1.80 ± 1.65	-0.27 ± 0.09
TYR CA	-1.22 ± 0.52	-0.86 ± 0.38	-0.41 ± 0.16	-1.70 ± 0.57	-0.27 ± 0.11
TYR CB	-1.75 ± 0.96	-0.90 ± 0.66	-0.56 ± 0.27	-2.28 ± 0.78	-0.49 ± 0.18
TYR CD1	-1.64 ± 0.85	-0.41 ± 0.47	-0.66 ± 0.33	-1.71 ± 0.77	-0.43 ± 0.19
TYR CD2	-1.67 ± 0.81	-0.42 ± 0.50	-0.62 ± 0.29	-1.53 ± 0.66	-0.38 ± 0.18
TYR CE1	-1.67 ± 0.94	-0.39 ± 0.45	-0.42 ± 0.26	-1.13 ± 0.66	-0.35 ± 0.19
TYR CE2	-1.70 ± 0.90	-0.39 ± 0.46	-0.41 ± 0.26	-1.10 ± 0.64	-0.34 ± 0.19
TYR CG	-0.90 ± 0.41	-0.23 ± 0.26	-0.33 ± 0.14	-1.12 ± 0.32	-0.21 ± 0.09
TYR CZ	-1.31 ± 0.61	-0.21 ± 0.28	-0.27 ± 0.15	-0.74 ± 0.38	-0.23 ± 0.13
TYR N	-1.30 ± 0.54	-1.06 ± 0.54	-0.79 ± 1.63	-2.59 ± 32.70	-0.36 ± 0.12
TYR O	-2.49 ± 0.93	-1.00 ± 0.62	-1.10 ± 0.45	-2.36 ± 1.11	-0.51 ± 0.28
TYR OH	-1.78 ± 1.17	-0.48 ± 0.52	-0.44 ± 0.35	-1.64 ± 1.36	-0.32 ± 0.28
VAL C	-1.58 ± 0.38	-0.19 ± 0.20	-0.24 ± 0.10	-1.70 ± 0.60	-0.27 ± 0.08
VAL CA	-1.45 ± 0.50	-0.21 ± 0.24	-0.39 ± 0.12	-1.70 ± 0.53	-0.27 ± 0.10
VAL CB	-1.39 ± 0.68	-0.30 ± 0.38	-0.40 ± 0.17	-1.57 ± 0.49	-0.40 ± 0.12
VAL CG1	-3.15 ± 1.57	-0.60 ± 0.82	-1.14 ± 0.37	-3.05 ± 0.92	-0.54 ± 0.25
VAL CG2	-3.22 ± 1.50	-0.62 ± 0.83	-1.02 ± 0.40	-2.93 ± 1.04	-0.54 ± 0.25
VAL N	-2.09 ± 0.55	-0.25 ± 0.28	-0.71 ± 0.33	-2.17 ± 0.81	-0.36 ± 0.10
VAL O	-3.33 ± 0.94	-0.44 ± 0.39	-1.08 ± 0.43	-2.32 ± 1.03	-0.53 ± 0.27

All units in kcal/mol. Errors represent the standard deviation and those values where the standard deviation was larger than the mean were considered meaningless for this analysis.

Lennard-Jones Repulsive Interactions

The Lennard-Jones Repulsive term reports heavily on very close range interactions, providing a different picture than the Lennard-Jones Attractive term. As was discussed in above, the repulsive term is only an approximation of the repulsive force between atoms and the magnitude grows too quickly when atoms are closely packed. Equation 10 calculates the repulsive term with an R^{-12} dependence on distance, which works well for distances on the order of the sum of the van der Waals radii. When atoms are closer, the repulsive term is more accurately represented by an equation of the form $e^{-\alpha R}$, where α is a function of the polarizability of the two atoms. In these calculations, there is an additional error introduced with the addition of the hydrogen atoms. The

crystallographic data does not provide precise locations for the hydrogen atoms, so their location must be approximated from the local bond geometries. Due to the inaccuracies of the repulsive term, the calculated values do not accurately reflect the forces in the molecule, but the relative magnitude of different groups can provide additional insight into the role different atom groups play in stabilizing folded proteins.

Figure 10 shows the total Lennard-Jones repulsive term normalized by protein size (number of heavy atoms). Stacking interaction between aromatic rings results in short distances between aromatic groups and a strong repulsive term. There also appears to be strong interactions of aromatic groups with aliphatic and polar groups. Due to the constraints placed by the backbone geometry, there is not a significant interaction between alpha carbons, as was seen with the Lennard-Jones attractive term.

Figure 11 shows the average Lennard-Jones Repulsive interaction for each amino acid. There are strong interactions between the charged groups, with the exception of lysine, and the aromatic groups. Cation-π interactions would lead to an increase in the lysine and arginine interactions with aromatic groups, but interestingly there is not a strong interaction between the aromatic groups and the lysine residue. The strong interactions by aspartic acid and glutamic acid suggest favorable interactions with positively charged histidine. Tyrosine and histidine have strong interactions with polar groups, consistent with the strong interactions seen between arginine, aspartic acid and glutamic acid and the aromatic groups.

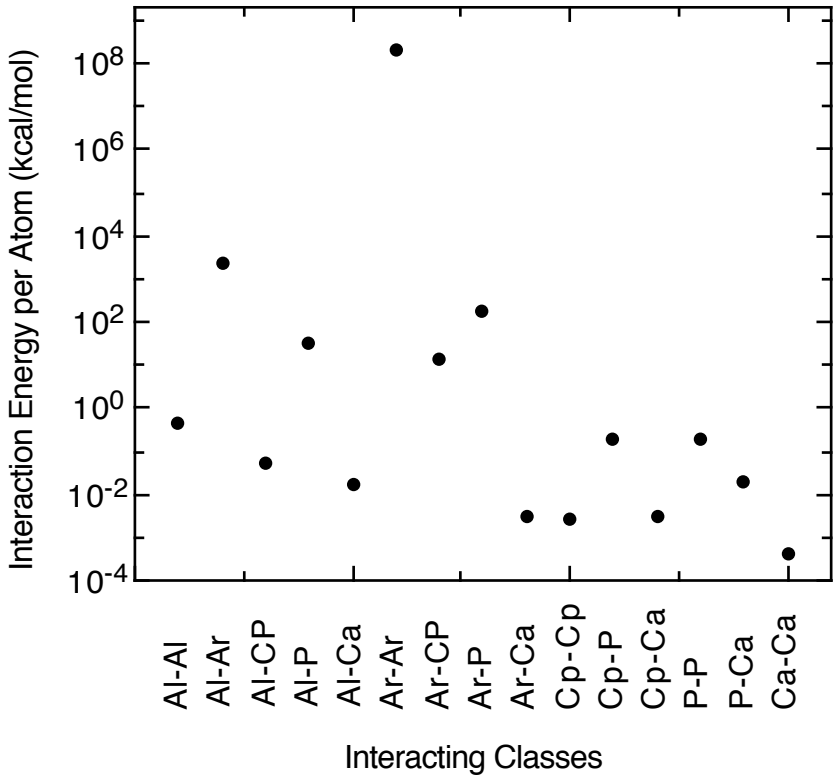


Figure 10: Lennard-Jones repulsive interactions divided into types. The interaction energy has been normalized for the size (number of atoms) of the protein.

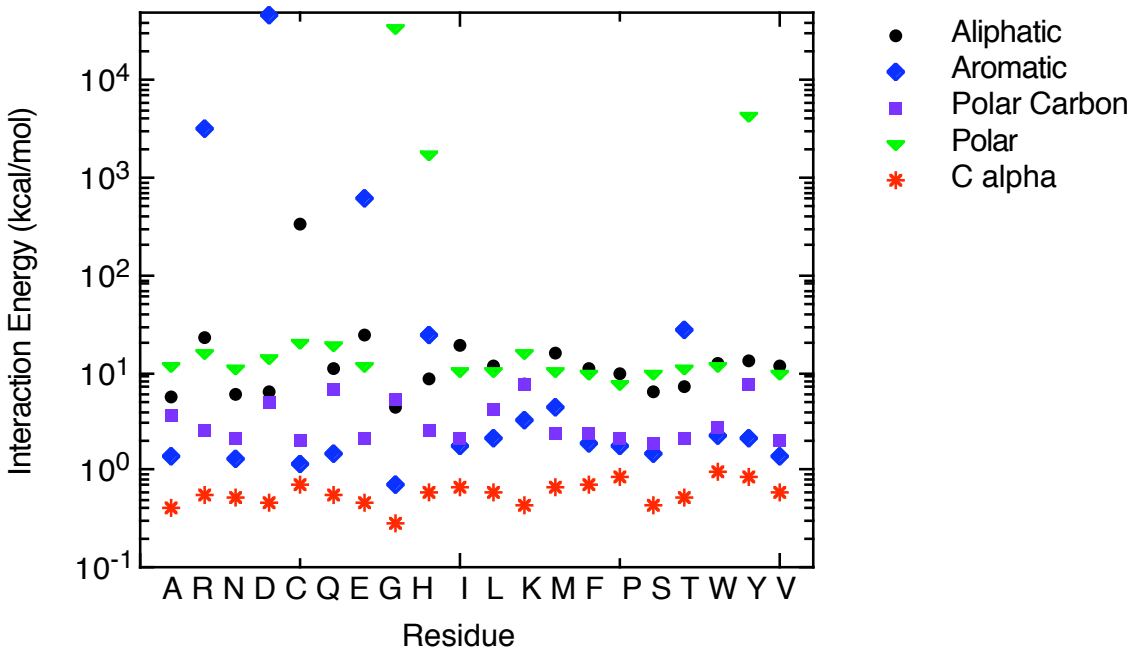


Figure 11: Average Lennard-Jones repulsive interaction between each residue and other atom groups in the protein.

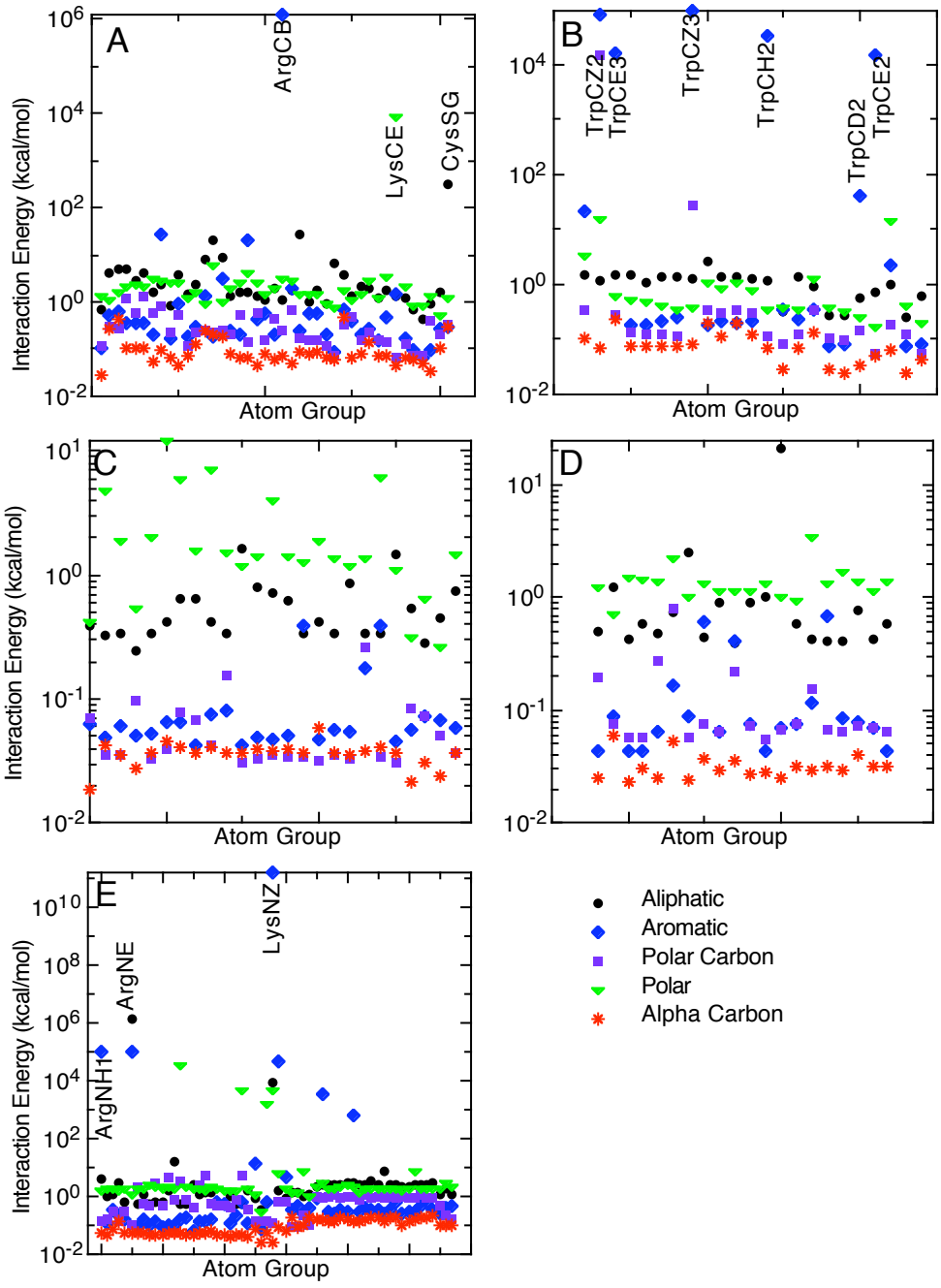


Figure 12: Lennard-Jones repulsive interactions between individual groups and all other groups. A) Aliphatic Groups B) Aromatic Groups C) Polar Carbons D) Alpha Carbon E) Polar Groups. Groups in each panel have been sorted by the van der Waals well depth parameter of AMBER 99.

Table 11: Lennard-Jones repulsive interactions between given atom group and class

Residue		Aliphatic	Aromatic	Polar Carbon	Polar	Alpha Carbon
ALA	C	0.32 ± 0.44	0.05 ± 0.15	0.04 ± 0.05	4.74 ± 296.14	0.04 ± 0.24
ALA	CA	0.58 ± 10.62	0.08 ± 0.33	0.08 ± 0.21	0.95 ± 1.07	0.03 ± 0.08
ALA	CB	2.34 ± 41.69	0.28 ± 0.97	0.23 ± 0.86	1.65 ± 3.69	0.12 ± 0.20
ALA	N	0.56 ± 0.82	0.09 ± 0.39	2.34 ± 239.71	1.83 ± 1.46	0.06 ± 0.17
ALA	O	2.12 ± 2.97	0.92 ± 82.41	0.98 ± 0.79	2.37 ± 43.63	0.16 ± 0.18
ARG	C	0.74 ± 1.83	0.06 ± 0.30	0.04 ± 0.06	1.45 ± 1.09	0.04 ± 0.02
ARG	CA	0.92 ± 4.54	0.07 ± 1.07	0.07 ± 0.12	1.13 ± 1.25	0.03 ± 0.07
ARG	CB	1.13 ± 16.88	1.3E+6 ± 1.3E+08	0.24 ± 0.36	3.24 ± 111.26	0.07 ± 0.08
ARG	CD	2.00 ± 51.89	0.27 ± 1.22	0.22 ± 1.70	2.78 ± 10.24	0.14 ± 0.41
ARG	CG	1.70 ± 63.45	1.93 ± 101.75	0.68 ± 0.78	2.93 ± 10.68	0.05 ± 0.10
ARG	CZ	0.45 ± 0.67	0.07 ± 0.66	0.05 ± 0.09	0.26 ± 0.37	0.02 ± 0.05
ARG	N	1.13 ± 3.04	0.16 ± 2.94	0.54 ± 0.49	1.83 ± 1.54	0.05 ± 0.03
ARG	NE	1.3E+6 ± 1.3E+08	1.1E+05 ± 1.0E+07	0.10 ± 0.22	1.18 ± 2.89	0.06 ± 0.14
ARG	NH1	4.34 ± 111.08	1.1E+05 ± 1.0E+07	0.14 ± 0.30	1.71 ± 9.49	0.06 ± 0.16
ARG	NH2	1.02 ± 20.39	0.12 ± 1.33	0.16 ± 0.29	1.84 ± 3.84	0.05 ± 0.12
ARG	O	2.70 ± 7.26	3209.00 ± 3.1E+05	0.95 ± 0.87	3.07 ± 130.69	0.16 ± 0.22
ASN	C	0.34 ± 0.39	0.06 ± 0.15	0.03 ± 0.06	1.89 ± 1.42	0.03 ± 0.02
ASN	CA	0.41 ± 1.74	0.09 ± 1.09	0.06 ± 0.16	1.68 ± 1.25	0.03 ± 0.12
ASN	CB	0.81 ± 9.09	0.17 ± 0.75	0.22 ± 0.46	2.53 ± 22.68	0.06 ± 0.08
ASN	CG	0.24 ± 0.44	0.05 ± 0.16	0.10 ± 0.11	0.55 ± 0.55	0.03 ± 0.04
ASN	N	0.55 ± 0.93	0.10 ± 0.54	0.51 ± 0.49	2.40 ± 1.84	0.05 ± 0.04
ASN	NE2	1.13 ± 5.15	0.36 ± 18.00	0.32 ± 1.01	1.99 ± 10.66	0.08 ± 0.16
ASN	O	2.44 ± 2.54	0.32 ± 3.73	1.08 ± 0.87	1.78 ± 5.59	0.14 ± 0.19
ASN	OE1	1.27 ± 3.24	0.26 ± 0.93	0.51 ± 0.81	1.76 ± 6.24	0.10 ± 0.17
ASP	C	0.33 ± 0.71	0.05 ± 0.13	0.03 ± 0.06	2.06 ± 1.26	0.04 ± 0.02
ASP	CA	0.42 ± 1.57	0.04 ± 0.20	0.06 ± 0.22	1.49 ± 4.00	0.02 ± 0.08
ASP	CB	0.71 ± 8.70	0.10 ± 0.59	0.12 ± 0.24	1.34 ± 2.65	0.03 ± 0.06
ASP	CG	0.28 ± 0.62	0.07 ± 1.22	0.07 ± 0.26	0.63 ± 1.02	0.03 ± 0.05
ASP	N	0.60 ± 2.35	0.14 ± 2.34	2.97 ± 271.97	2.34 ± 2.23	0.05 ± 0.21
ASP	O	2.54 ± 14.67	0.28 ± 0.91	1.11 ± 0.89	1.84 ± 1.99	0.14 ± 0.18
ASP	OD1	1.66 ± 26.60	4.6E+04 ± 5.1E+06	0.63 ± 0.95	6.31 ± 403.85	0.09 ± 0.49
ASP	OD2	0.83 ± 3.05	4.59 ± 458.44	0.65 ± 48.30	1.86 ± 3.90	0.07 ± 0.15
CYS	C	0.42 ± 1.54	0.06 ± 0.17	0.04 ± 0.07	12.23 ± 592.17	0.05 ± 0.42
CYS	CA	0.78 ± 1.11	0.08 ± 0.37	0.07 ± 0.09	1.40 ± 1.30	0.04 ± 0.08
CYS	CB	27.64 ± 1249.59	0.24 ± 0.55	0.17 ± 0.27	1.41 ± 1.31	0.09 ± 0.10
CYS	N	0.66 ± 1.44	0.09 ± 0.20	0.45 ± 0.45	1.80 ± 1.36	0.05 ± 0.03
CYS	O	2.48 ± 3.63	0.37 ± 1.66	0.89 ± 0.79	2.13 ± 17.49	0.19 ± 0.22
CYS	SG	309.80 ± 409.45	0.30 ± 0.67	0.32 ± 0.36	1.21 ± 1.24	0.30 ± 0.35
GLN	C	0.63 ± 0.77	0.06 ± 1.32	0.08 ± 0.15	5.94 ± 379.48	0.04 ± 0.28
GLN	CA	0.47 ± 2.07	0.06 ± 0.94	0.28 ± 2.57	1.41 ± 2.29	0.03 ± 0.07
GLN	CB	1.72 ± 72.76	0.46 ± 21.44	0.14 ± 0.41	3.40 ± 87.55	0.07 ± 0.33
GLN	CD	0.54 ± 8.63	0.06 ± 0.18	0.08 ± 0.15	0.31 ± 0.52	0.02 ± 0.04
GLN	CG	1.56 ± 52.20	0.21 ± 1.39	0.57 ± 0.77	3.14 ± 20.42	0.05 ± 0.11
GLN	N	1.72 ± 56.35	0.10 ± 0.70	5.05 ± 372.59	2.15 ± 12.02	0.06 ± 0.37
GLN	NE2	3.16 ± 130.95	0.23 ± 3.10	0.14 ± 0.30	1.61 ± 15.97	0.15 ± 0.30
GLN	O	2.82 ± 10.94	0.29 ± 1.47	1.03 ± 1.55	2.51 ± 49.37	0.16 ± 0.45
GLN	OE1	1.21 ± 3.66	0.40 ± 13.64	0.10 ± 0.29	0.97 ± 3.23	0.18 ± 0.60
GLU	C	0.64 ± 0.89	0.04 ± 0.12	0.07 ± 0.13	1.57 ± 1.12	0.04 ± 0.02
GLU	CA	0.51 ± 5.38	0.04 ± 0.24	0.19 ± 0.44	1.27 ± 2.60	0.02 ± 0.09
GLU	CB	1.18 ± 42.16	0.17 ± 3.91	0.12 ± 0.54	2.08 ± 5.00	0.06 ± 0.20
GLU	CD	0.39 ± 0.68	0.06 ± 0.23	0.07 ± 0.14	0.41 ± 0.69	0.02 ± 0.04
GLU	CG	3.79 ± 159.40	0.89 ± 78.06	0.54 ± 0.92	2.67 ± 18.35	0.05 ± 0.12
GLU	N	15.01 ± 1575.53	0.10 ± 1.16	0.72 ± 0.77	1.97 ± 1.86	0.05 ± 0.03
GLU	O	2.98 ± 32.72	613.40 ± 6.9E+04	0.98 ± 0.82	2.11 ± 49.85	0.13 ± 0.17
GLU	OE1	1.30 ± 4.06	0.33 ± 2.58	0.09 ± 0.24	1.32 ± 3.19	0.18 ± 0.39
GLU	OE2	1.36 ± 63.02	0.45 ± 8.13	0.08 ± 0.25	1.33 ± 3.36	0.09 ± 0.21

Table 11: Continued

Residue		Aliphatic	Aromatic	Polar Carbon	Polar	Alpha Carbon
GLY	C	0.41 ± 0.65	0.07 ± 0.23	0.04 ± 0.08	6.99 ± 431.56	0.04 ± 0.27
GLY	CA	0.74 ± 2.63	0.17 ± 0.63	0.80 ± 0.46	2.20 ± 11.08	0.05 ± 0.11
GLY	N	0.58 ± 2.89	0.18 ± 3.30	3.70 ± 287.11	3.5E+0 ± 4.4E+06	0.06 ± 0.28
GLY	O	2.58 ± 75.40	0.31 ± 2.15	0.81 ± 0.77	1.44 ± 1.72	0.13 ± 0.20
HIS	C	0.34 ± 0.39	0.08 ± 0.21	0.15 ± 0.23	1.50 ± 1.15	0.04 ± 0.02
HIS	CA	0.42 ± 0.99	0.12 ± 0.31	0.16 ± 0.28	3.41 ± 138.96	0.03 ± 0.09
HIS	CB	1.28 ± 19.89	0.25 ± 0.72	0.20 ± 0.31	1.90 ± 2.58	0.08 ± 0.10
HIS	CD2	1.51 ± 44.64	20.69 ± 746.40	0.35 ± 0.91	3.33 ± 15.20	0.10 ± 0.15
HIS	CE1	0.94 ± 9.66	2.09 ± 86.72	0.18 ± 0.86	13.80 ± 813.16	0.06 ± 0.11
HIS	CG	0.25 ± 0.36	0.07 ± 0.19	0.12 ± 0.15	0.41 ± 0.39	0.02 ± 0.04
HIS	N	0.57 ± 0.95	0.19 ± 3.37	0.71 ± 0.59	2.08 ± 1.51	0.05 ± 0.03
HIS	ND1	0.65 ± 1.00	0.17 ± 0.60	0.31 ± 0.48	1.78 ± 2.24	0.05 ± 0.10
HIS	NE2	0.62 ± 1.53	0.66 ± 24.61	0.14 ± 0.29	1664.00 ± 1.1E+05	0.05 ± 0.09
HIS	O	2.47 ± 3.53	0.38 ± 1.86	1.12 ± 0.91	1.86 ± 1.95	0.17 ± 0.20
ILE	C	1.64 ± 0.77	0.04 ± 0.11	0.03 ± 0.04	1.18 ± 0.97	0.04 ± 0.02
ILE	CA	1.01 ± 1.67	0.04 ± 0.19	0.06 ± 0.07	1.34 ± 1.08	0.03 ± 0.06
ILE	CB	6.83 ± 634.87	0.09 ± 1.04	0.07 ± 0.12	0.78 ± 1.17	0.06 ± 0.06
ILE	CD1	8.32 ± 215.56	1.31 ± 76.46	0.24 ± 0.68	0.94 ± 4.59	0.24 ± 0.37
ILE	CG1	1.56 ± 11.85	0.20 ± 0.82	0.54 ± 0.74	2.46 ± 10.38	0.07 ± 0.10
ILE	CG2	4.14 ± 20.58	0.34 ± 2.11	1.28 ± 0.63	2.09 ± 2.53	0.10 ± 0.16
ILE	N	2.67 ± 77.59	0.07 ± 0.20	0.39 ± 0.39	1.90 ± 1.63	0.05 ± 0.03
ILE	O	3.24 ± 2.82	0.38 ± 1.37	0.90 ± 0.81	1.95 ± 1.55	0.20 ± 0.21
LEU	C	0.79 ± 1.08	0.05 ± 0.11	0.03 ± 0.05	1.42 ± 0.99	0.04 ± 0.02
LEU	CA	2.48 ± 7.26	0.09 ± 5.34	0.06 ± 0.19	1.04 ± 1.08	0.02 ± 0.05
LEU	CB	1.40 ± 4.18	0.18 ± 1.40	0.12 ± 0.22	1.23 ± 2.04	0.07 ± 0.10
LEU	CD1	3.97 ± 49.74	0.53 ± 9.64	0.30 ± 0.64	1.12 ± 25.96	0.28 ± 0.34
LEU	CD2	5.08 ± 68.06	0.64 ± 24.94	0.28 ± 0.72	1.59 ± 89.12	0.45 ± 0.49
LEU	CG	0.90 ± 44.01	0.10 ± 0.37	0.39 ± 0.68	1.29 ± 1.95	0.04 ± 0.06
LEU	N	1.26 ± 1.45	0.15 ± 4.09	2.46 ± 260.43	2.03 ± 1.45	0.05 ± 0.17
LEU	O	2.76 ± 6.80	0.36 ± 1.12	1.00 ± 0.82	1.97 ± 1.77	0.18 ± 0.20
LYS	C	0.73 ± 2.88	0.05 ± 0.13	0.04 ± 0.07	4.00 ± 275.95	0.04 ± 0.20
LYS	CA	21.16 ± 2016.15	0.07 ± 1.26	0.07 ± 0.30	1.00 ± 1.56	0.02 ± 0.08
LYS	CB	1.63 ± 27.68	19.65 ± 2023.18	0.14 ± 0.23	4.01 ± 224.87	0.06 ± 0.11
LYS	CD	21.34 ± 2015.41	0.19 ± 3.07	0.22 ± 2.83	6.10 ± 428.65	0.20 ± 1.04
LYS	CE	1.38 ± 32.19	1.40 ± 23.13	0.07 ± 0.17	8314.00 ± 6043.91	0.04 ± 0.21
LYS	CG	1.37 ± 34.81	0.43 ± 27.98	0.56 ± 0.82	2.48 ± 8.45	0.05 ± 0.13
LYS	N	1.35 ± 7.59	0.13 ± 3.74	5.80 ± 403.54	1.75 ± 1.54	0.06 ± 0.34
LYS	NZ	8339.00 ± 6444.96	1.6E+11 ± 1.7E+14	0.13 ± 1.86	4468.00 ± 5435.12	0.03 ± 0.16
LYS	O	7.38 ± 433.70	0.26 ± 0.73	0.92 ± 0.89	1.93 ± 38.15	0.14 ± 0.19
MET	C	0.62 ± 0.85	0.05 ± 0.12	0.03 ± 0.05	1.40 ± 1.08	0.04 ± 0.02
MET	CA	0.92 ± 1.50	0.06 ± 0.25	0.06 ± 0.09	1.13 ± 1.79	0.03 ± 0.09
MET	CB	2.07 ± 10.20	0.19 ± 0.65	0.15 ± 0.48	1.45 ± 2.26	0.08 ± 0.09
MET	CE	8.63 ± 93.37	3.19 ± 157.23	0.24 ± 0.73	0.98 ± 2.75	0.19 ± 0.49
MET	CG	1.99 ± 56.03	0.20 ± 0.62	0.42 ± 0.77	1.95 ± 5.69	0.06 ± 0.10
MET	N	1.33 ± 11.51	0.18 ± 3.38	0.59 ± 0.51	2.23 ± 1.59	0.06 ± 0.04
MET	O	2.75 ± 5.98	0.35 ± 1.01	0.98 ± 0.80	2.02 ± 8.77	0.18 ± 0.19
MET	SD	1.66 ± 1.98	0.26 ± 0.63	0.20 ± 0.45	0.53 ± 2.17	0.10 ± 0.15

Table 11: Continued

Residue		Aliphatic	Aromatic	Polar Carbon	Polar	Alpha Carbon
PHE	C	0.34 ± 0.51	0.38 ± 0.49	0.03 ± 0.06	1.27 ± 1.00	0.04 ± 0.02
PHE	CA	0.41 ± 0.93	0.68 ± 1.19	0.07 ± 0.16	1.36 ± 4.78	0.03 ± 0.09
PHE	CB	1.05 ± 2.39	0.55 ± 1.23	0.16 ± 0.25	1.47 ± 1.29	0.08 ± 0.09
PHE	CD1	1.29 ± 7.07	0.19 ± 0.93	0.33 ± 0.49	1.05 ± 1.88	0.20 ± 0.28
PHE	CD2	1.23 ± 10.79	0.20 ± 0.78	0.29 ± 0.37	0.74 ± 1.10	0.11 ± 0.19
PHE	CE1	1.38 ± 11.91	0.22 ± 0.49	0.12 ± 0.22	0.38 ± 1.42	0.08 ± 0.10
PHE	CE2	1.30 ± 4.62	0.26 ± 2.08	0.11 ± 0.30	0.35 ± 0.91	0.07 ± 0.10
PHE	CG	0.27 ± 0.31	0.08 ± 0.17	0.09 ± 0.11	0.32 ± 0.28	0.02 ± 0.03
PHE	CZ	1.39 ± 13.84	0.23 ± 0.67	0.12 ± 0.28	0.34 ± 0.85	0.07 ± 0.11
PHE	N	0.57 ± 1.64	0.68 ± 2.01	0.46 ± 0.45	1.85 ± 1.35	0.05 ± 0.03
PHE	O	2.44 ± 3.31	0.63 ± 0.99	0.91 ± 0.79	1.91 ± 1.64	0.20 ± 0.22
PRO	C	0.41 ± 0.43	0.05 ± 0.12	0.03 ± 0.06	1.92 ± 1.02	0.06 ± 0.13
PRO	CA	1.25 ± 63.54	0.09 ± 0.39	0.07 ± 0.22	0.71 ± 0.92	0.06 ± 0.18
PRO	CB	0.92 ± 6.95	0.19 ± 0.70	0.11 ± 0.30	0.88 ± 1.60	0.06 ± 0.13
PRO	CD	3.76 ± 38.07	0.71 ± 50.40	0.31 ± 0.61	1.81 ± 3.01	0.46 ± 0.25
PRO	CG	1.35 ± 39.49	0.38 ± 14.36	0.45 ± 7.40	1.09 ± 24.28	0.07 ± 0.10
PRO	N	0.38 ± 0.46	0.07 ± 0.16	0.13 ± 0.17	0.31 ± 0.43	0.03 ± 0.02
PRO	O	2.38 ± 4.10	0.28 ± 1.23	1.06 ± 0.91	1.54 ± 2.80	0.11 ± 0.16
SER	C	0.34 ± 0.71	0.06 ± 0.14	0.03 ± 0.06	1.37 ± 0.97	0.04 ± 0.04
SER	CA	0.43 ± 2.06	0.07 ± 0.41	0.07 ± 0.16	1.16 ± 1.32	0.03 ± 0.09
SER	CB	1.16 ± 33.22	0.16 ± 0.58	0.15 ± 0.28	1.22 ± 1.87	0.07 ± 0.09
SER	N	0.56 ± 1.24	0.56 ± 47.71	0.48 ± 0.49	1.67 ± 8.03	0.05 ± 0.03
SER	O	2.14 ± 3.82	0.30 ± 0.76	0.91 ± 0.84	1.88 ± 1.95	0.15 ± 0.19
SER	OG	2.01 ± 106.54	0.39 ± 8.35	0.22 ± 0.42	2.89 ± 4.34	0.11 ± 0.18
THR	C	0.85 ± 0.66	0.05 ± 0.14	0.03 ± 0.06	1.17 ± 0.90	0.04 ± 0.02
THR	CA	0.59 ± 8.35	0.04 ± 0.22	0.06 ± 0.12	1.42 ± 1.51	0.03 ± 0.08
THR	CB	0.44 ± 3.44	0.07 ± 0.29	0.07 ± 0.18	0.85 ± 1.18	0.05 ± 0.05
THR	CG2	2.34 ± 33.03	26.52 ± 2002.66	0.84 ± 0.67	2.90 ± 28.76	0.10 ± 0.21
THR	N	1.10 ± 4.37	0.11 ± 2.02	0.41 ± 0.45	1.59 ± 1.46	0.04 ± 0.04
THR	O	2.72 ± 6.55	0.32 ± 1.16	0.93 ± 0.85	7.08 ± 560.05	0.17 ± 0.20
THR	OG1	1.01 ± 4.37	0.25 ± 2.09	0.22 ± 0.30	7.73 ± 560.04	0.10 ± 0.16
TRP	C	0.34 ± 0.41	0.18 ± 0.37	0.26 ± 0.27	1.36 ± 1.02	0.04 ± 0.02
TRP	CA	0.40 ± 1.30	0.41 ± 0.62	0.22 ± 0.20	1.15 ± 1.01	0.04 ± 0.11
TRP	CB	1.08 ± 2.34	0.58 ± 0.63	0.16 ± 0.22	1.41 ± 1.32	0.08 ± 0.08
TRP	CD1	0.92 ± 2.55	0.34 ± 0.39	0.33 ± 0.40	1.21 ± 1.54	0.12 ± 0.15
TRP	CD2	0.54 ± 0.57	39.29 ± 2.63	0.14 ± 0.17	0.25 ± 0.24	0.03 ± 0.05
TRP	CE2	0.69 ± 0.93	1.5E+04 ± 1331.88	0.05 ± 0.08	0.16 ± 0.20	0.05 ± 0.07
TRP	CE3	1.45 ± 1.92	1.6E+04 ± 1790.09	0.26 ± 0.43	0.60 ± 0.94	0.22 ± 0.19
TRP	CG	0.30 ± 0.35	0.33 ± 0.19	0.08 ± 0.10	0.36 ± 0.31	0.03 ± 0.04
TRP	CH2	1.11 ± 1.67	3.2E+04 ± 2159.33	0.11 ± 0.19	0.34 ± 0.81	0.07 ± 0.12
TRP	CZ2	1.17 ± 3.41	8.1E+04 ± 4.8E+04	1.4E+04 ± 1332.63	14.87 ± 1.74	0.07 ± 0.12
TRP	CZ3	1.28 ± 3.72	9.2E+04 ± 4.8E+04	27.07 ± 2.12	0.37 ± 0.70	0.08 ± 0.12
TRP	N	0.57 ± 1.04	0.22 ± 0.34	0.81 ± 0.58	1.75 ± 1.30	0.05 ± 0.03
TRP	NE1	0.89 ± 1.96	14.69 ± 2.60	0.14 ± 0.20	1.18 ± 1.46	0.08 ± 0.10
TRP	O	2.50 ± 7.58	0.48 ± 0.78	1.14 ± 0.91	1.99 ± 1.63	0.19 ± 0.20

Table 11: Continued

Residue		Aliphatic			Aromatic			Polar Carbon			Polar			Alpha Carbon		
TYR	C	0.34	±	0.36	0.39	±	0.48	0.03	±	0.05	6.13	±	410.93	0.04	±	0.27
TYR	CA	0.45	±	2.41	0.61	±	1.84	0.08	±	0.32	1.33	±	1.20	0.04	±	0.11
TYR	CB	1.74	±	55.21	0.54	±	0.85	0.16	±	0.22	1.50	±	1.43	0.08	±	0.10
TYR	CD1	2.54	±	118.03	0.18	±	0.41	0.35	±	0.46	1.09	±	1.49	0.19	±	0.24
TYR	CD2	1.33	±	12.21	0.21	±	2.20	0.30	±	0.37	0.82	±	1.42	0.11	±	0.17
TYR	CE1	1.45	±	29.95	0.17	±	0.53	0.13	±	0.56	0.50	±	2.20	0.07	±	0.11
TYR	CE2	1.05	±	2.63	0.17	±	0.62	0.12	±	0.22	0.48	±	0.94	0.07	±	0.12
TYR	CG	0.27	±	0.34	0.07	±	0.19	0.10	±	0.12	0.36	±	0.35	0.03	±	0.05
TYR	CZ	0.62	±	1.18	0.08	±	0.21	0.05	±	0.09	0.20	±	0.25	0.04	±	0.07
TYR	N	0.57	±	1.38	0.70	±	1.34	5.71	±	439.60	4327.00	±	3.6E+05	0.05	±	0.37
TYR	O	2.49	±	4.81	0.64	±	1.40	0.90	±	0.80	1.92	±	1.76	0.20	±	0.23
TYR	OH	1.29	±	2.53	0.46	±	6.01	0.18	±	0.34	2.09	±	4.30	0.10	±	0.19
VAL	C	1.46	±	0.57	0.05	±	0.12	0.03	±	0.04	1.10	±	0.93	0.04	±	0.02
VAL	CA	0.58	±	2.42	0.04	±	0.19	0.06	±	0.10	1.43	±	1.21	0.03	±	0.06
VAL	CB	0.69	±	8.43	0.10	±	0.33	0.07	±	0.13	0.81	±	1.19	0.06	±	0.06
VAL	CG1	4.93	±	194.97	0.36	±	2.24	1.19	±	0.63	2.12	±	2.76	0.11	±	0.17
VAL	CG2	2.89	±	18.75	0.37	±	1.74	0.57	±	0.71	2.39	±	2.88	0.11	±	0.19
VAL	N	1.55	±	0.92	0.12	±	5.42	0.36	±	0.37	1.83	±	1.35	0.04	±	0.03
VAL	O	3.22	±	9.32	0.37	±	0.72	0.84	±	0.78	1.90	±	1.61	0.21	±	0.22

All units in kcal/mol. Errors represent the standard deviation and those values where the standard deviation was larger than the mean were considered meaningless for this analysis.

Figure 12 shows the average contribution of individual atom groups to the overall Lennard-Jones repulsive interactions of the protein. Table 11 shows the data used to make Figure 12. Since the interaction depends on the well depth, the groups in each panel have been sorted by the well depth. In each classification, a few atom groups appear significantly different. In panel A, lysine Cε and cysteine Sγ are seen as outliers for the same reasons discussed above. Arginine Cβ also stands out but has a very large deviation and should not be considered significant. In panel B, tryptophan dominates the interactions with other aromatic groups. In panel E, lysine Nζ again stands out with a significant error, but a few more groups appear to be interesting. Arginine Nε and Nη1 both have significant interactions with aromatic groups due to cation-π interactions.

Conclusion

This study has looked at the average non-bonded interactions of various atom groups in folded proteins. From the data some general trends are obvious. The electrostatic interactions with polar groups have the opposite sign from the electrostatic interactions with other groups, an effect of polar groups tending to have a partial negative charge with other groups tending to have a partial positive charge. There is also an ordering in the strength of electrostatic interactions; from strongest to weakest: polar carbon, aromatic groups, aliphatic groups and alpha carbons. Similar trends can be seen for the van der Waals interactions, with polar and aliphatic groups generally having larger interactions than polar carbons, aromatic groups, and alpha carbons. Table 12 provides a summary of the interactions that are significantly different from these trends. Looking at the interactions of tryptophan, the strong van der waals interactions of the six member ring indicate aromatic stacking interactions occur primarily in this area. There are also interesting electrostatic interactions on the Cβ side of the ring structure. Looking at the interactions of arginine, the Nη groups have weak electrostatic interactions but the Nε has strong Lennard-Jones interactions with aliphatic groups, suggesting the head of the side chain is solvent exposed, but the body packs with aliphatic chains in the protein interior.

Table 12: Summary of significant interactions

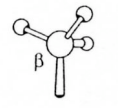
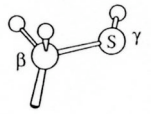
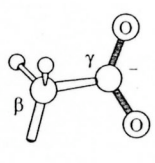
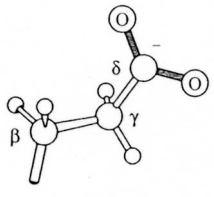
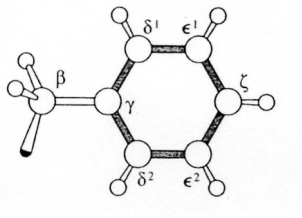
 ALA	N CA C O CB
 CYS	N CA C O CB SG Strong R to Cα
 ASP	N CA C O CB CG OD1 OD2
 GLU	N CA C O CB CG CD OE1 OE2
 PHE	N CA C O CB CG CD1 CD2 CE1 CE2 CZ

Table 12: Continued

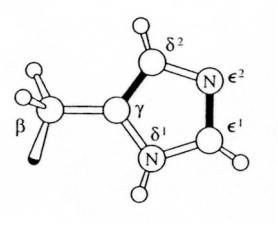
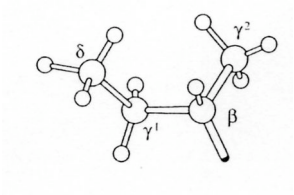
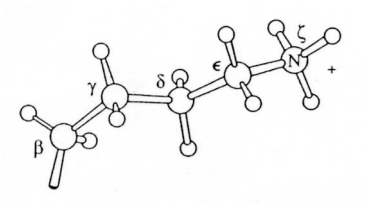
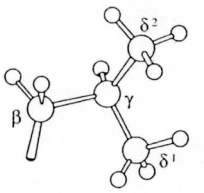
 GLY	N	Strong R to polar groups
	CA	Strong E and A
	C	
	O	
 HIS	N	
	CA	
	C	
	O	
	CB	
	CG	
	ND1	
	CD2	
	CE1	
	NE2	Strong R to polar groups
 ILE	N	
	CA	
	C	
	O	
	CB	
	CG1	
	CG2	
	CD1	
 LYS	N	
	CA	
	C	
	O	
	CB	
	CG	
	CD	
	CE	Strong A,R to polar groups
	NZ	Strong E, strong A, R to aromatic and polar groups
 LEU	N	
	CA	
	C	
	O	
	CB	
	CG	
	CD1	Strong R to Cα
	CD2	Strong R to Cα

Table 12: Continued

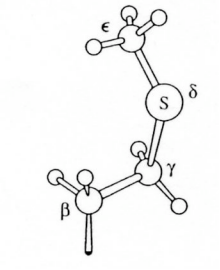
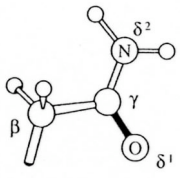
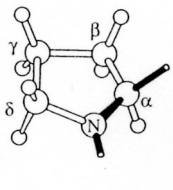
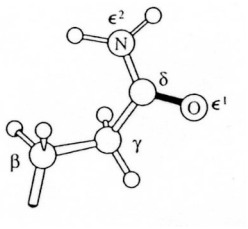
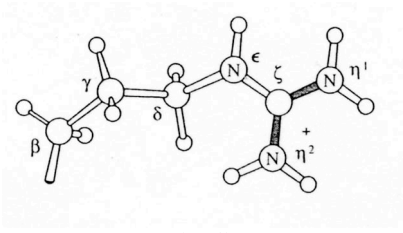
 MET	N	
	CA	
	C	
	O	
	CB	
	CG	Strong E
	SD	Strong E
	CE	Strong E
 ASN	N	
	CA	
	C	
	O	
	CB	
	CG	
	OD1	
	ND2	Weak E, strong R to Cα
 PRO	N	Strong E
	CA	
	C	
	O	
	CB	
	CG	
	CD	Strong R to Cα
 GLN	N	
	CA	
	C	
	O	
	CB	
	CG	
	CD	
	OE1	
	NE2	Weak E, strong R to Cα
 ARG	N	
	CA	
	C	
	O	
	CB	
	CG	
	CD	
	NE	Strong R to Aliphatic
	CZ	
	NH1	Weak E
	NH2	Weak E

Table 12: Continued

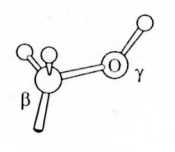
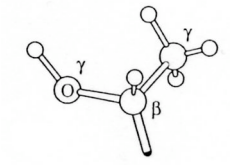
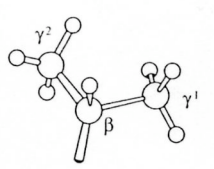
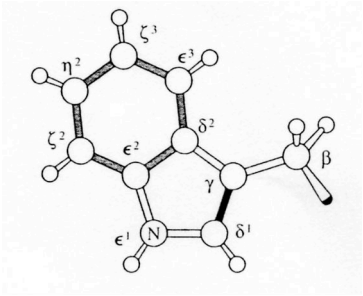
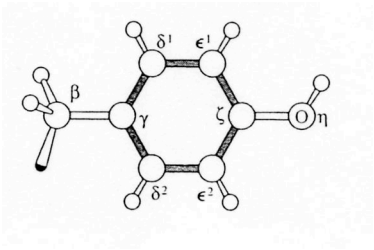
 SER	N CA C O CB OG
 THR	N CA C O CB OG1 CG2
 VAL	N CA C O CB CG1 CG2
 TRP	N CA C O CB CG Strong E CD1 NE1 CD2 Strong A to aromatic, strong R to polar carbon CE2 Strong E, strong A to aromatic CE3 Strong A to aromatic and polar carbons CZ2 Strong A to aromatic and polar carbons CZ3 Strong A to aromatic CH2 Strong A to aromatic

Table 12: Continued

 TYR	N CA C O CB CG CD1 CD2 CE1 CE2 CZ OH	Strong R to polar

Strong/weak is relative to other atom groups in the same class with similar charge or well depth. E, A, R refer to electrostatic, Lennard-Jones attractive and Lennard-Jones repulsive terms. Side chain models from (Creighton 1997).

PACKING OF BURIED ATOMS

Introduction

When a protein folds to its native three-dimensional (3D) structure, the packing of atoms exceeds the packing density of close packed spheres (Richards 1974; Chothia 1975). The increased packing makes the van der Waals forces stronger in the native state than in the denatured state. Studies of leucine to alanine mutations in T4 Lysozyme have shown a relationship between increases in cavity size and the loss of protein stability (Matthews et al. 1987). Kono et al showed that cavity-filling mutations contribute 2 to 3 kcal/mol per methylene group (Kono et al. 2000).

While packing in proteins approaches that of close packed spheres, the interior of the protein is not uniform. Kuntz found distinct hydrophobic and polar regions in carboxypeptidase, noting that polar regions had a higher density (g/cc) than hydrophobic regions (Kuntz 1972). Tsai et al note that when surface waters are included in the calculation, the protein surface is as well packed as the interior (Tsai et al. 1999). Others have shown that small cavities and other packing irregularities are common in proteins (Hubbard et al. 1994; Pontius et al. 1996; Tsai et al. 2002). It is important to understand what features give rise to the packing differences.

Finney defined packing density as the ratio of the volume of the sphere described by the van der Waals radius and the total volume assigned to the group in the structure (Finney

1970). In order to calculate packing density, the volume occupied by the atom group must be determined. Since it is difficult to determine the location of hydrogen atoms using conventional crystallographic techniques, hydrogen atoms are generally not reported in crystal structures. To account for the unknown location of the hydrogen, I utilize a unified atom set where the volume of the hydrogen is included in the volume of the atom to which it is covalently attached. Voronoi described a method of assigning space to fixed points (Voronoi 1908). The Voronoi volume for an atom group contains all points that are closer to the group center than any other group center. Richards described improvements to incorporate the differences in atomic radii between atom groups (Richards 1974).

In this study, I investigated the features that give rise to packing irregularities. Honig suggested hydrogen bonds might act to stabilize proteins by increasing the local packing (Honig 1999). To look at this, I have investigated the difference that hydrogen bonding makes to packing. Others have shown that the depth of burial can also affect packing (Hubbard et al. 1994). Therefore, I investigated changes in packing between surface, boundary and core atoms. I also looked at how secondary structure alters packing.

Methods

Packing Densities

The local packing density is defined as the ratio between the volume of a sphere with a radius equal to the van der Waals radius for the atom and the total volume occupied by that atom (Finney 1970).

13 $$PD_{loc} = \frac{V_{vdW}}{V_{tot}}$$

Using the atomic radii determined by Chothia shown in Table 13 (Chothia 1975), I calculated the total volume using Richards' Method B (Richards 1974) implemented in the code-mbg library (Harpaz et al. 1994, Gerstein, 1992 #406; Gerstein et al. 1995).

Protein Data Set

I used PISCES (Wang and Dunbrack 2002) to generate a non-redundant set of high-resolution protein structures. The criteria I used to select structures were: 1) X-ray structures with resolution better than 1.8 Å, 2) less than 50% identical and 3) have between 80 and 1000 amino acids. From this set, I eliminated all structures that contained gaps and missing atoms, which results in 872 structures for analysis. Table 14 contains the pdb codes and the chain identifiers.

Table 13: Chothia radii

Atom Type	Radii	Protein atoms
C4/CPro	1.87	All_CA,ALA_CB,ARG_CB,ARG_CG,ARG_CD,ASN_CB,ASP_CB,CYS_CB, GLN_CB,GLN_CG,GLU_CB,GLU_CG,HIS_CB,ILE_CB,ILE_CG1,ILE_CG2, ILE_CD,LEU_CB,LEU_CG,LEU_CD1,LEU_CD2,LYS_CB,LYS_CG,LYS_CD, LYS_CE,MET_CB,MET_CG,MET_CE,PHE_CB,PRO_CB,PRO_CG,PRO_CD, SER_CB,THR_CB,THR_CG2,TRP_CB,TYR_CB,VAL_CB,VAL_CG1, VAL_CG2
C3	1.76	All_C,ARG_CZ,ASN_CG,ASP_CG,GLN_CD,GLU_CD,HIS_CG,HIS_CD2, HIS_CE1,PHE_CG,PHE_CD1,PHE_CD2,PHE_CE1,PHE_CE2,PHE_CZ, TRP_CG,TRP_CD1,TRP_CD2,TRP_CE2,TRP_CE3,TRP_CZ2,TRP_CZ3, TRP_CH2,TYR_CG,TYR_CD1,TYR_CD2,TYR_CE1,TYR_CE2,TYR_CZ
O	1.40	All_O,ASN_OD1,ASP_OD1,ASP_OD2,GLN_OE1,GLU_OE1, GLU_OE2, SER_OG,THR_OG1,TYR_OH
N4/N3	1.5	HIS_ND1, LYS_NZ
N3H	1.65	All_N,ARG_NE,ARG_NH1,ARG_NH2,ASN_ND2,GLN_NE2,HIS_NE2, TRP_NE1
S	1.85	CYS_SG,MET_SD

Table 14: Protein data set (872 Files)

119l	A	1clx	A	1eca	A	1g5t	A	1hm9	A	1jg1	A	1lam	A	1nar	A	1tfe	A
16pk	A	1cmc	A	1ecs	A	1g61	A	1hmt	A	1jh6	A	1lb6	A	1nb9	A	1thf	D
19hc	A	1cnv	A	1ed8	A	1g66	A	1hn2	A	1jhd	A	1lbu	A	1nbc	A	1thm	A
1a12	A	1cnz	A	1edg	A	1g6g	A	1hnj	A	1jhf	A	1lbv	A	1nep	A	1thx	A
1a1i	A	1co6	A	1eeo	A	1g6h	A	1hqk	A	1jhg	A	1lc0	A	1nff	A	1tml	A
1a28	A	1cot	A	1ehd	A	1g6s	A	1htr	B	1jhj	A	1lc5	A	1nfp	A	1toa	A
1a2p	A	1cpq	A	1ej0	A	1g8e	A	1htw	A	1jid	A	1ld8	A	1nkp	A	1tx4	A
1a3a	A	1cqm	A	1ej8	A	1g8k	B	1hw1	A	1jig	A	1ld8	B	1nkp	B	1uah	A
1a6m	A	1cqx	A	1ek6	A	1g8q	A	1hx0	A	1jiw	I	1lf2	A	1nkr	A	1udh	A
1a73	A	1cs6	A	1ekg	A	1g94	A	1hx6	A	1jix	A	1lfw	A	1nlb	H	1ugi	A
1a8d	A	1csh	A	1el5	A	1g9o	A	1hxh	A	1jjf	A	1lj5	A	1nlq	A	1uro	A
1a8e	A	1ctj	A	1elk	A	1g9z	A	1hyo	A	1jjt	A	1lj9	A	1nls	A	1ute	A
1a8q	A	1ctq	A	1elu	A	1ga6	A	1hyp	A	1jk3	A	1lk5	A	1nme	A	1uxy	A
1aba	A	1cuo	A	1elw	A	1gbs	A	1hz4	A	1jke	A	1lko	A	1nme	B	1vca	A
1ads	A	1cv8	A	1emv	A	1gca	A	1hzt	A	1jkv	A	1llp	A	1noa	A	1vfr	A
1ag9	A	1cxc	A	1emv	B	1gco	A	1i0d	A	1jkx	A	1lmb	3	1nox	A	1vhh	A
1agi	A	1cxy	A	1enf	A	1gd0	A	1i0r	A	1jl1	A	1lmi	A	1npk	A	1vsr	A
1agj	A	1cyd	A	1eok	A	1gde	A	1i0v	A	1jlj	A	1lni	A	1nsc	A	1wad	A
1ah7	A	1cyo	A	1ep0	A	1gdo	A	1i1j	A	1jlv	A	1lo7	A	1nsz	A	1wba	A
1ajs	A	1czf	A	1epx	A	1gdv	A	1i1n	A	1jm1	A	1lok	A	1nth	A	1wer	A
1ak0	A	1czp	A	1eqo	A	1geg	A	1i1w	A	1jnr	B	1lop	A	1nu4	A	1whi	A
1ako	A	1d02	A	1erv	A	1gg6	B	1i40	A	1jo0	A	1lpl	A	1nwz	A	1xgs	A
1aky	A	1d0c	A	1erz	A	1gg6	C	1i4u	A	1jq5	A	1lq9	A	1nyt	A	1xnb	A
1al3	A	1d0q	A	1es5	A	1giq	A	1i52	A	1jr8	A	1lqp	A	1nza	A	1xyz	A
1amf	A	1d1q	A	1es9	A	1gk8	I	1i58	A	1jsd	A	1lqv	A	1nzy	A	1yac	A
1amm	A	1d2n	A	1euv	A	1gk9	A	1i5g	A	1jsd	B	1lri	A	1o08	A	1yna	A
1aoe	A	1d2v	A	1euv	B	1gkl	A	1i5r	A	1jta	A	1lst	A	1o1z	A	1zin	A
1aoh	A	1d3v	A	1euw	A	1gmu	A	1i71	A	1jtg	B	1luc	B	1o7j	A	256b	A
1aqb	A	1d4a	A	1evh	A	1gnu	A	1i7h	A	1jv4	A	1ly2	A	1o7n	B	2a0b	A
1arb	A	1d4o	A	1evl	A	1gny	A	1i8f	A	1jvw	A	1lyc	A	1oa2	A	2act	A
1atg	A	1d4t	A	1ew0	A	1go2	A	1i8o	A	1jx6	A	1lyv	A	1oaa	A	2acy	A
1atl	A	1d4x	G	1ew4	A	1go3	F	1iab	A	1jye	A	1lzl	A	1oaf	A	2ahj	B
1atz	A	1d7p	M	1ew6	A	1goi	A	1iby	A	1jyh	A	1m07	A	1oal	A	2apr	A
1auo	A	1dbf	A	1ewf	A	1gp0	A	1ic6	A	1jyr	A	1m0u	A	1ock	A	2arc	A
1axn	A	1dbo	A	1exm	A	1gp6	A	1icr	A	1jzg	A	1m15	A	1onc	A	2ayh	A
1ay7	B	1dbw	A	1exr	A	1gpi	A	1id0	A	1k0i	A	1m1n	A	1one	A	2baa	A
1ayx	A	1dci	A	1ey4	A	1gpq	A	1ida	A	1k0m	A	1m1q	A	1ooe	A	2bbk	H
1b0b	A	1df7	A	1eyh	A	1gq8	A	1iej	A	1k1e	A	1m26	A	1opd	A	2bop	A
1b16	A	1dfu	P	1eyv	A	1gqa	A	1ifc	A	1k20	A	1m2d	A	1pch	A	2bvw	A
1b2p	A	1dgf	A	1ezg	A	1gqv	A	1ifr	A	1k2e	A	1m3k	A	1pdo	A	2ccy	A
1b5e	A	1dgw	A	1ezm	A	1gs5	A	1ift	A	1k2y	X	1m4i	A	1pgs	A	2cpl	A
1b8o	A	1dgw	Y	1ezw	A	1gt9	1	1ig0	A	1k3y	A	1m4j	A	1pgt	A	2cth	A
1b9o	A	1dhn	A	1f0y	A	1gtv	A	1ihj	A	1k4g	A	1m4l	A	1pgx	A	2cua	A
1b9w	A	1dj0	A	1f1m	A	1gtz	A	1iho	A	1k4i	A	1m55	A	1php	A	2cy3	A
1bbh	A	1dj7	A	1f2t	B	1gu2	A	1iib	A	1k5c	A	1m5e	A	1plc	A	2eif	A
1bd0	A	1dk0	A	1f3u	A	1gu7	A	1ijb	A	1k5n	A	1m7g	A	1pmi	A	2end	A
1bd8	A	1dk8	A	1f46	A	1gud	A	1ijt	A	1k5n	B	1m7j	A	1pot	A	2fcb	A
1bdo	A	1dl5	A	1f4p	A	1guq	A	1ijy	A	1k6k	A	1m7s	A	1ppn	A	2fcr	A
1beb	A	1dlf	L	1f5v	A	1gve	A	1ikt	A	1k6w	A	1m9z	A	1psr	A	2gdm	A
1beh	A	1dlj	A	1f60	A	1gvo	A	1im5	A	1k7c	A	1mba	A	1ptf	A	2hmz	A
1bf6	A	1dlw	A	1f60	B	1gvp	A	1io0	A	1k7i	A	1mfa	H	1qau	A	2hvm	A
1bfg	A	1dly	A	1f74	A	1gwe	A	1io7	A	1k92	A	1mfg	A	1qaz	A	2ilk	A
1bg2	A	1dmh	A	1f7d	A	1gwm	A	1ioo	A	1k94	A	1mfm	A	1qb7	A	2lis	A
1bgf	A	1doz	A	1f7l	A	1gx3	A	1iq4	A	1ka1	A	1mgt	A	1qcx	A	2ltn	A
1bj7	A	1dps	A	1f86	A	1gx4	A	1iq6	A	1kaf	A	1mh9	A	1qd9	A	2mcm	A
1bk0	A	1dpt	A	1f8m	A	1gxm	A	1iqc	A	1kao	A	1mix	A	1qdd	A	2mhr	A

Table 14: Continued

1bk7	A	1dqe	A	1fao	A	1gxu	A	1iqq	A	1kdj	A	1mjn	A	1qf9	A	2nac	A
1bkf	A	1dqg	A	1faz	A	1gxy	A	1iqz	A	1kep	A	1mk0	A	1qft	A	2nlr	A
1bkp	A	1dqi	A	1fcy	A	1gy6	A	1ird	A	1kew	A	1mkk	A	1qgi	A	2por	A
1bkr	A	1dqp	A	1fec	A	1gy7	A	1ird	B	1kfw	A	1ml4	A	1qh4	A	2pth	A
1bm8	A	1dqz	A	1fh0	A	1gyo	A	1ire	B	1khi	A	1mla	A	1qh5	A	2rhe	A
1bn7	A	1ds1	A	1fh9	A	1gyv	A	1is3	A	1kid	A	1mml	A	1qhq	A	2sak	A
1bn8	A	1dsz	A	1fi2	A	1gzc	A	1isp	A	1km4	A	1mn8	A	1qhv	A	2sic	I
1bqb	A	1dug	A	1fiu	A	1gzg	A	1it2	A	1kmt	A	1mol	A	1qip	A	2spc	A
1bqc	A	1duv	G	1fk5	A	1gzt	A	1itx	A	1kng	A	1mop	A	1qj4	A	2tgi	A
1bqk	A	1dxe	A	1fl0	A	1h03	P	1iu8	A	1knm	A	1moq	A	1qjc	A	2tps	A
1brt	A	1dym	A	1flm	A	1h0h	B	1iua	A	1koe	A	1mpg	A	1qkk	A	2trx	A
1bsm	A	1dyq	A	1flt	X	1h2e	A	1iup	A	1koi	A	1mqk	H	1ql0	A	3bam	A
1bup	A	1dys	A	1fm0	D	1h2r	S	1iv3	A	1kol	A	1mqv	A	1ql3	A	3c2c	A
1bx4	A	1dz3	A	1fmb	A	1h4g	A	1ix9	A	1kpf	A	1mr3	F	1qlw	A	3cao	A
1bxa	A	1dzk	A	1fmc	A	1h4r	A	1ixh	A	1kpt	A	1mrj	A	1qmq	A	3chb	D
1byi	A	1dzo	A	1fn9	A	1h5q	A	1iz7	A	1kq3	A	1msk	A	1qna	A	3cyr	A
1byq	A	1e0w	A	1fna	A	1h6f	A	1izc	A	1kqf	B	1mty	B	1qnf	A	3dfr	A
1c02	A	1e12	A	1fnd	A	1h6h	A	1j09	A	1kqf	C	1mty	G	1qnn	A	3eip	A
1c0p	A	1e19	A	1fnl	A	1h6l	A	1j1x	H	1kqp	A	1mtz	A	1qnr	A	3ezm	A
1c1d	A	1e1a	A	1fp2	A	1h6t	A	1j54	A	1kqr	A	1mug	A	1qop	B	3grs	A
1c1k	A	1e29	A	1fpo	A	1h6u	A	1j5r	A	1kqw	A	1mun	A	1qq4	A	3lzt	A
1c1l	A	1e2w	A	1fqt	A	1h70	A	1j6n	A	1kr7	A	1muw	A	1qq5	A	3pvi	A
1c24	A	1e30	A	1fs5	A	1h72	C	1j71	A	1krh	A	1mvo	A	1qqf	A	3sdh	A
1c3p	A	1e42	A	1fs7	A	1h75	A	1j8u	A	1ks8	A	1mwp	A	1qre	A	3seb	A
1c44	A	1e43	A	1ft5	A	1h7n	A	1j96	A	1ks9	A	1mxi	A	1qrr	A	3sil	A
1c52	A	1e4c	P	1ftr	A	1h7z	A	1j9q	A	1ktg	A	1mxr	A	1qs1	A	3vub	A
1c5e	A	1e4m	M	1fvk	A	1h8u	A	1ja9	A	1kuf	A	1my7	A	1qsg	A	451c	A
1c7k	A	1e58	A	1fvu	A	1h97	A	1jak	A	1kv8	A	1n08	A	1qst	A	4eug	A
1cbs	A	1e5k	A	1fw9	A	1h9m	A	1jat	A	1kw3	B	1n13	B	1qtn	A	4fiv	A
1ccw	A	1e5m	A	1fx2	A	1h9o	A	1jay	A	1kwf	A	1n1j	A	1qtn	B	4pga	A
1ccw	B	1e6c	A	1fxl	A	1hbn	B	1jb3	A	1kwn	A	1n1j	B	1qto	A	4ubp	A
1ccz	A	1e6u	A	1fxo	A	1hbn	C	1jcl	A	1kyf	A	1n3y	A	1qtw	A	4ubp	B
1cex	A	1e7l	A	1fzq	A	1hd2	A	1jd1	A	1kzk	A	1n45	A	1ra9	A	5hpg	A
1cg5	A	1e85	A	1g0o	A	1hdi	A	1jdr	A	1kzq	A	1n55	A	1rcf	A	5nul	A
1cg5	B	1e87	A	1g12	A	1hdk	A	1jer	A	1l5o	A	1n62	A	1rie	A	5pal	A
1chd	A	1e9g	A	1g1t	A	1hdo	A	1jf2	A	1l6p	A	1n62	C	1rro	A	6gsv	A
1ci9	A	1ea7	A	1g2a	A	1hfe	S	1jf3	A	1l6r	A	1n71	A	1sbp	A	7a3h	A
1cip	A	1eaj	A	1g2o	A	1hfo	A	1jf8	A	1l6x	A	1n83	A	1sml	A	7fd1	A
1cjc	A	1ear	A	1g2q	A	1hg8	A	1jfb	A	1l7a	A	1n8k	A	1swu	A	8abp	A
1cjw	A	1eaz	A	1g2r	A	1hh8	A	1jfu	A	1l7l	A	1n8v	A	1t1d	A	8tln	E
1cl8	A	1eb6	A	1g4i	A	1hm6	A	1jfx	A	1l9x	A	1n97	A	1tca	A		

Four letter PDB code and chain identifier.

Classification

Atoms were classified based on burial, hydrogen bonding and secondary structure. Figure 13 shows how atom burial was divided into three groups, surface (atoms that contact the molecular surface), boundary (atoms that contact surface atoms, but not the molecular surface) and core (atoms that do not contact the molecular surface or surface atoms). The boundary and core classes were combined to form a buried class to investigate the extent to which packing is affected by hydrogen bonding and secondary structure. Amino acids were divided into two groups, surface (amino acids with at least one surface atom group) and buried (amino acids with no surface atom groups). To determine the classification of atom groups and amino acids, the molecular surface and atomic contacts were calculated using the code-mbg library (Harpaz et al. 1994, Gerstein, 1992 #406; Gerstein et al. 1995), hydrogen bonding was determined using HBPLUS (McDonald and Thornton 1994) and secondary structure was determined using DSSP (Kabsch and Sander 1983). DSSP utilizes hydrogen bonding patterns to determine secondary structure. Hydrogen bonds are classified as n-turns if the hydrogen bond forms between CO_i and NH_{i+1} for n=3,4,5 or bridges if they are distant in sequence. Alpha helix is defined as repeating 4-turns while beta structure is defined as repeating bridges.

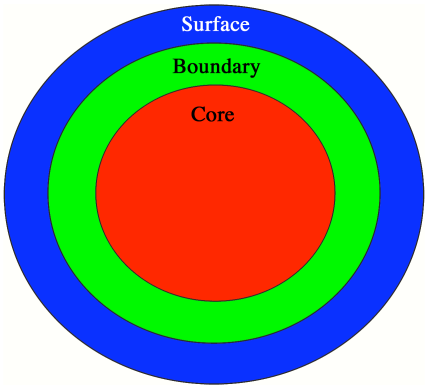


Figure 13: Atom Burial Classification

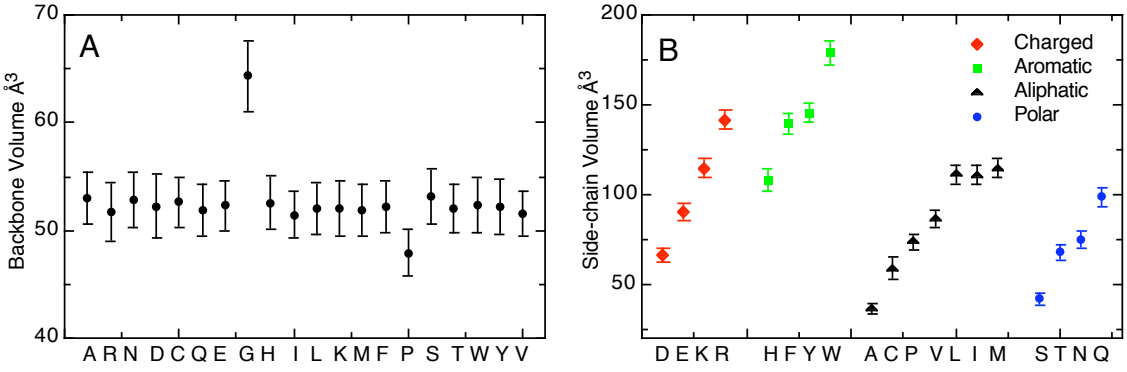


Figure 14: Mean volumes for each buried residue. A) Backbone volumes B) Side chain volumes. Error bars represent one standard deviation from the mean.

Results

Amino Acid Volumes

To understand how hydrogen bonding and secondary structure affect packing in a protein, I first looked at how these affect the volumes occupied by amino acids in the folded protein. Since most amino acids contain an identical backbone and a variable side chain, I separated the volumes of the backbone from the side chain. Figure 14A shows the measured volumes for the backbones of all 20 amino acids. As expected, the volumes of the backbone are similar (about 52.2 $Å^3$) across the majority of amino acids. There are two notable exceptions, glycine with an average volume of 64.3 $Å^3$ (12.1 $Å^3$ larger) and proline with an average volume of 48.0 $Å^3$ (4.2 $Å^3$ smaller). Unlike most other amino acids, glycine has a single hydrogen atom for a side chain instead of a beta carbon attached to the alpha carbon. Since I am including the attached hydrogen atoms in the volume of the alpha carbon, the volume of the glycine backbone is increased, to a small extent, by the volume of the second hydrogen. In addition, in other amino acids, the beta carbon is covalently attached to the alpha carbon. The van der Waals radii of the alpha and beta carbon are approximately 1.8 Å whereas the bond distance is about 1.5 Å. Without the covalent bond, the closest approach would be about 3.6 Å. This difference, combined with the addition of the hydrogen atom, result in a significant increase in the volume of the alpha carbon for glycine. Proline has a decreased volume for similar reasons. The side chain of proline forms a 5-member ring that includes the backbone

nitrogen and the alpha carbon. The covalent bond of the Cδ and the loss of the hydrogen atom reduces the size of the backbone nitrogen. Figure 14B shows the average volumes of the amino acid side-chains. As has been previously seen, longer side chains have larger volumes (Richards 1974; Chothia 1975; Harpaz et al. 1994; Pontius et al. 1996; Tsai and Gerstein 2002).

Table 15: Differences in residue volume associated with solvent exposure

Amino Acid	Backbone		Side Chain		Total	
	Buried Volume ($Å^3$)	Surface ΔV ($Å^3$)	Buried Volume ($Å^3$)	Surface ΔV ($Å^3$)	Buried Volume ($Å^3$)	Surface ΔV ($Å^3$)
ASP	52.25	8.13	66.38	23.99	118.63	40.36
GLU	52.34	6.94	90.29	27.58	142.63	45.75
LYS	52.03	7.55	114.50	30.35	166.53	58.55
ARG	51.71	8.30	141.60	35.72	193.31	61.30
HIS	52.60	6.93	107.90	26.30	160.50	40.53
PHE	52.17	7.94	139.80	18.80	191.97	27.14
TYR	52.19	6.62	145.50	24.60	197.69	36.62
TRP	52.33	6.50	179.00	25.62	231.33	35.40
GLY	64.31	24.06	NA	NA	64.31	24.06
ALA	53.00	11.43	36.16	12.71	89.16	21.84
CYS	52.62	11.18	58.93	14.41	111.55	24.67
PRO	47.95	11.28	73.67	21.28	121.62	33.32
VAL	51.63	7.58	86.39	15.08	138.02	21.04
LEU	51.99	7.63	111.30	17.17	163.29	24.23
ILE	51.42	6.73	111.00	16.13	162.42	21.43
MET	51.92	7.66	114.50	20.59	166.42	27.36
SER	53.11	10.36	42.06	17.46	95.17	28.64
THR	52.04	8.45	67.95	21.18	119.99	30.84
ASN	52.81	7.64	74.77	23.95	127.58	37.37
GLN	51.93	7.00	98.90	28.25	150.83	43.10

$\Delta V = V_{Surface} - V_{Buried}$; NA for "not applicable"

Table 15 shows how the volume changes when residues are on the protein surface. These differences are an artifact of the calculation, but demonstrate why the surface residues were excluded from further analysis. No surface water was included in the calculation, creating two known problems with atoms at the surface of the protein. If the atom is significantly exposed to the solvent, the calculation results in an infinite volume as there are missing edges to the Voronoi polyhedra. These atoms are excluded from all sets. In the other case, the infinite polyhedra of some atoms provide closing edges for other atoms. In this case, a volume is determined that is larger than the volume that the atom would occupy in solution. Tsai et al has shown when water is included in the calculation, the volumes determined are consistent with buried residues (Tsai et al. 1999). Since I am principally interested in packing, ignoring these surface artifacts presents no real problems.

Table 16 shows the impact of hydrogen bonding on the volume of buried amino acids. On average, there is a decrease (1.4 $Å^3$) between the backbone volume of amino acids that have no backbone hydrogen bonds and those where either the backbone nitrogen or the carbonyl forms a hydrogen bond. For amino acids where both the backbone nitrogen and carbonyl form hydrogen bonds there is a larger decease (3.5 $Å^3$) in the backbone volume. With the exception of histidine, there is also a decrease in the side chain volume associated with the formation of hydrogen bonds.

Table 16: Differences in volume associated with hydrogen bonding in buried residues

	Backbone			Side chain		
	No H Bonds	1 H Bond	2 H Bonds	No H Bonds	1+ H Bonds	
Residue	Volume (Å^3)	ΔV (Å^3)	ΔV (Å^3)	Volume (Å^3)	Volume (Å^3)	ΔV (Å^3)
ALA	57.2 (3.8 , 22)	-2.7 (2.8 , 228)	-4.5 (2.3 , 1890)			
ARG	55.9 (NA , 1)	-0.4 (3.0 , 3)	-5.2 (1.7 , 16)	NF	141.6 (5.3 , 20)	ND
ASN	56.5 (2.3 , 5)	-2.5 (2.7 , 26)	-4.0 (2.4 , 140)	78.0 (5.6 , 10)	74.6 (4.5 , 161)	-3.4
ASP	64.0 (NA , 1)	-9.8 (2.4 , 26)	-12.4 (2.6 , 95)	71.8 (4.6 , 2)	66.3 (4.0 , 120)	-5.5
CYS	54.5 (2.4 , 5)	-0.2 (2.7 , 48)	-2.1 (2.1 , 329)			
GLN	53.5 (NA , 1)	0.4 (2.6 , 9)	-1.8 (2.3 , 73)	100.4 (4.6 , 8)	98.7 (5.3 , 75)	-1.7
GLU	NF	ND	ND	92.6 (6.8 , 4)	90.2 (4.5 , 62)	-2.4
GLY	66.4 (3.4 , 28)	-0.7 (3.6 , 259)	-2.5 (3.1 , 1088)			
HIS	56.8 (1.3 , 2)	-2.9 (2.2 , 16)	-4.6 (2.3 , 68)	102.7 (0.6 , 2)	108.0 (6.1 , 84)	5.3
ILE	54.5 (3.2 , 14)	-1.1 (3.1 , 99)	-3.2 (2.0 , 1324)			
LEU	55.9 (3.9 , 8)	-1.5 (2.8 , 156)	-4.1 (2.2 , 1593)			
LYS	NF	ND	ND	NF	114.5 (5.2 , 19)	ND
MET	57.0 (3.0 , 4)	-3.6 (2.6 , 42)	-5.4 (2.2 , 357)			
PHE	55.2 (0.3 , 4)	-0.8 (2.7 , 61)	-3.2 (2.2 , 590)			
PRO	49.2 (2.2 , 34)	-1.4 (2.0 , 175)	NA			
SER	55.9 (3.9 , 13)	-1.2 (2.5 , 92)	-3.2 (2.4 , 426)	45.5 (4.7 , 23)	41.9 (3.2 , 508)	-3.6
THR	53.7 (3.4 , 12)	-0.6 (2.3 , 78)	-1.9 (2.2 , 399)	69.6 (5.4 , 31)	67.8 (4.1 , 458)	-1.7
TRP	NF	ND	ND	179.8 (6.5 , 19)	178.9 (6.8 , 80)	-0.9
TYR	55.7 (5.3 , 2)	-1.5 (2.8 , 22)	-3.8 (2.3 , 173)	149.2 (5.1 , 22)	145.0 (5.3 , 175)	-4.2
VAL	54.6 (3.0 , 11)	-1.2 (3.0 , 164)	-3.2 (1.9 , 1741)			
Average		-1.4[1] (2.8 , 1530)	-3.5[1] (2.2 , 10460)			

$\Delta V = V_{Hbond} - V_{No\ H\ Bonds}$; NA for "not applicable"; NF for "not found"; ND for "not determined".
Values in parentheses are the standard deviation and the number of observations.
[1] Aspartic acid was excluded from the average.

Table 17 shows the differences in the backbone and side chain volumes associated with different secondary structures for buried amino acids. The alpha helix is a more compact backbone configuration with an increased side chain volume. In general, secondary structures that increase the volume of the backbone appear to decrease the volume of the side chains. The only exception to this is the beta strand. This type of secondary structure is more expanded for both the backbone and side chain. Random coil has a significant decrease in the side chain volume and a minor increase in the backbone. Without the restricted conformation of a regular secondary structure element, the side chains are able to significantly improve the packing without a large disruption in the packing of the backbone. Favorable packing in random coil regions could have a significant impact on the overall stability of the folded protein.

Atom Groups

Table 18 shows the effect of solvent exposure on the local packing density (PD_{loc}) of atom groups. The differences in packing between the core and boundary atom groups is small (±0.06), while the surface atoms show a significant reduction in local packing density ($-0.7 < \Delta PD < -0.2$). The difference seen in the surface atom groups is an artifact of the calculation due to the absence of surface waters to pack against, as has been discussed above. Due to the similarities of the boundary and core atom groups, these were combined to form a buried class to look at hydrogen bonding and secondary structure.

Table 17: Volume changes associated with differences in buried secondary structures

Secondary Structure Type	Backbone ΔV (Å^3)	Side Chain ΔV (Å^3)
α-Helix (repeating 4-turn)	-0.84 (5036)	0.12 (4093)
Isolated β-Bridge	0.18 (142)	-1.31 (123)
β-Strand (repeating β-bridge)	0.64 (5200)	0.46 (4664)
3_{10} Helix (repeating 3-turn)	0.56 (270)	-1.03 (228)
Isolated H Bonded Turn	1.05 (338)	-0.43 (262)
Bend (high curvature, no H bond)	0.85 (280)	-0.40 (185)
Random Coil[1]	0.35 (890)	-1.72 (726)

$\Delta V = Volume_{Secondary\ Structure} - Volume_{Buried}$; Value in parethesis is the number of observations.

[1] No structure was assigned by DSSP

Table 19 shows the effect of hydrogen bonding and secondary structure on the packing of atom groups. Consistent with the observation that hydrogen bonded amino acids have a decrease in volume, the packing of atom groups is increased when hydrogen bonds are formed. It is interesting that the hydrogen bonding of the backbone oxygen appears to increase the packing of the backbone carbonyl as well. The difference of about 0.03 is smaller than the standard deviation for the packing of the carbonyl, but is consistent across all amino acids and suggests that hydrogen bonding may improve packing of atoms not directly involved in the hydrogen bonding pair. The differences associated with different secondary structure are generally small. In an alpha helix, the volume of the backbone was reduced (-0.84 Å^3), but the packing of the Cα is reduced by about 0.08 (indicating a larger volume). While the overall volume is decreased due to increased packing of the C, N and O, it is interesting that the Cα is not well packed in the alpha helix.

Table 18: Packing of atom groups associated with solvent exposure

Residue		Core PD_{loc}	Boundary ΔPD	Surface ΔPD
GLY	C	2.32	-0.06	-0.51
	CA	1.18	0.00	-0.36
	N	1.30	-0.02	-0.40
	O	0.73	-0.01	-0.30
ALA	C	2.45	0.03	-0.18
	CA	1.97	-0.01	-0.59
	CB	0.77	-0.01	-0.18
	N	1.38	-0.01	-0.36
	O	0.72	0.01	-0.27
VAL	C	2.53	0.05	-0.26
	CA	2.14	-0.06	-0.52
	CB	1.91	-0.02	-0.46
	CG1	0.77	0.00	-0.16
	CG2	0.77	0.00	-0.15
	N	1.40	0.00	-0.42
	O	0.72	0.01	-0.26
LEU	C	2.49	0.01	-0.26
	CA	2.09	-0.01	-0.35
	CB	1.21	-0.01	-0.27
	CD1	0.75	0.00	-0.15
	CD2	0.75	0.00	-0.16
	CG	1.92	-0.01	-0.42
	N	1.38	0.03	-0.41
	O	0.72	0.02	-0.28
ILE	C	2.56	0.05	-0.27
	CA	2.14	-0.04	-0.48
	CB	1.96	0.00	-0.45
	CD1	0.73	0.01	-0.13
	CG1	1.16	0.01	-0.25
	CG2	0.78	0.00	-0.16
	N	1.39	0.02	-0.40
	O	0.71	0.03	-0.24
PRO	C	2.48	0.01	-0.18
	CA	2.00	-0.03	-0.65
	CB	1.09	0.02	-0.31
	CD	1.16	0.03	-0.33
	CG	1.08	0.01	-0.32
	N	2.20	-0.04	-0.48
	O	0.73	0.00	-0.31
CYS	C	2.44	0.05	-0.23
	CA	2.06	-0.02	-0.50
	CB	1.19	-0.02	-0.30
	N	1.34	0.04	-0.36
	O	0.73	0.00	-0.30
	SG	0.78	0.01	-0.20
SER	C	2.50	-0.04	-0.24
	CA	2.03	-0.02	-0.63
	CB	1.14	0.01	-0.31
	N	1.35	0.00	-0.36
	O	0.72	0.02	-0.27
	OG	0.64	0.01	-0.27
THR	C	2.51	0.02	-0.32
	CA	2.07	0.00	-0.57
	CB	1.91	0.00	-0.47
	CG2	0.78	-0.01	-0.20
	N	1.39	-0.01	-0.36
	O	0.74	0.00	-0.28
	OG1	0.67	0.01	-0.28
ASN	C	2.49	-0.01	-0.35
	CA	2.06	0.01	-0.50
	CB	1.18	0.01	-0.36
	CG	2.33	-0.02	-0.57
	N	1.38	0.02	-0.42
	ND2	0.78	0.03	-0.29
	O	0.71	0.03	-0.26
	OD1	0.68	0.03	-0.28
GLN	C	2.49	0.03	-0.26
	CA	2.11	-0.05	-0.58
	CB	1.19	0.01	-0.35
	CD	2.28	0.00	-0.56
	CG	1.19	-0.01	-0.36
	N	1.36	0.06	-0.37
	NE2	0.77	0.03	-0.29
	O	0.74	0.01	-0.28
	OE1	0.68	-0.01	-0.31
ASP	C	2.46	0.05	-0.24
	CA	2.07	-0.02	-0.53
	CB	1.20	-0.04	-0.40
	CG	2.36	-0.03	-0.59
	N	1.33	0.06	-0.37
	O	0.72	0.03	-0.27
	OD1	0.73	0.01	-0.33
	OD2	0.70	0.01	-0.33
GLU	C	2.49	0.04	-0.21
	CA	2.09	-0.04	-0.59
	CB	1.20	0.00	-0.39
	CD	2.29	0.02	-0.58
	CG	1.18	-0.01	-0.38
	N	1.36	0.05	-0.36
	O	0.73	0.01	-0.27
	OE1	0.75	-0.05	-0.39
	OE2	0.67	0.03	-0.31
MET	C	2.51	0.00	-0.28
	CA	2.08	-0.02	-0.52
	CB	1.19	0.00	-0.29
	CE	0.76	0.01	-0.18
	CG	1.19	-0.01	-0.33
	N	1.40	0.01	-0.40
	O	0.72	0.02	-0.27
	SD	0.86	0.00	-0.21

Table 18: Continued

Residue		Core PD_{loc}	Boundary ΔPD	Surface ΔPD
LYS	C	2.47	0.05	-0.22
	CA	2.08	-0.03	-0.58
	CB	1.14	0.05	-0.31
	CD	1.21	-0.06	-0.47
	CE	1.13	0.04	-0.37
	CG	1.13	0.05	-0.31
	N	1.39	0.01	-0.40
	NZ	0.61	0.10	-0.28
	O	0.72	0.02	-0.27
ARG	C	2.46	0.05	-0.22
	CA	2.07	-0.03	-0.57
	CB	1.18	0.02	-0.31
	CD	1.25	-0.04	-0.43
	CG	1.21	-0.03	-0.35
	CZ	2.30	0.00	-0.73
	N	1.39	0.01	-0.39
	NE	1.28	-0.06	-0.55
	NH1	0.84	-0.02	-0.35
	NH2	0.82	-0.02	-0.34
	O	0.72	0.02	-0.27
TRP	C	2.48	0.06	-0.28
	CA	2.07	-0.03	-0.50
	CB	1.16	0.00	-0.26
	CD1	1.07	0.04	-0.28
	CD2	2.15	0.01	-0.54
	CE2	2.21	0.00	-0.61
	CE3	1.10	0.02	-0.25
	CG	2.21	0.00	-0.58
	CH2	1.10	-0.02	-0.32
	CZ2	1.10	0.00	-0.33
	CZ3	1.06	0.01	-0.26
	N	1.38	0.02	-0.39
	NE1	1.13	-0.02	-0.45
	O	0.72	0.03	-0.26
PHE	C	2.50	0.03	-0.37
	CA	2.06	-0.02	-0.48
	CB	1.17	0.00	-0.24
	CD1	1.13	0.01	-0.28
	CD2	1.10	0.02	-0.26
	CE1	1.08	0.00	-0.29
	CE2	1.07	0.00	-0.28
	CG	2.28	-0.01	-0.59
	CZ	1.07	0.01	-0.27
	N	1.39	0.02	-0.38
	O	0.73	0.00	-0.28
TYR	C	2.48	0.05	-0.35
	CA	2.08	-0.03	-0.50
	CB	1.16	0.01	-0.24
	CD1	1.13	0.02	-0.29
	CD2	1.13	0.00	-0.29
	CE1	1.09	0.02	-0.30
	CE2	1.11	0.00	-0.32
	CG	2.28	-0.02	-0.59
	CZ	2.22	0.00	-0.49
	N	1.40	0.00	-0.40
	O	0.72	0.02	-0.27
	OH	0.63	0.01	-0.26
HIS	C	2.46	0.05	-0.32
	CA	2.07	-0.04	-0.53
	CB	1.16	0.01	-0.28
	CD2	1.08	0.04	-0.30
	CE1	1.14	-0.01	-0.41
	CG	2.24	0.00	-0.60
	N	1.38	0.02	-0.37
	ND1	1.01	-0.03	-0.44
	NE2	1.13	0.02	-0.47
	O	0.73	0.01	-0.28

$\Delta PD = PD_{other} - PD_{core}$

Table 19: Differences in packing due to hydrogen bonding and secondary structure

Residue		No H Bonds Packing	1 H Bond ΔPack	Average Packing	Beta Bridge ΔPack	Random Coil[1] ΔPack	Beta Ladder ΔPack	3/10 Helix ΔPack	Alpha Helix ΔPack	Bend ΔPack	Turn ΔPack
ALA	C	2.45	0.03[2]	2.48	-0.04	-0.04	-0.13	-0.02	0.08	-0.04	-0.03
	CA			1.96	0.08	0.07	0.09	-0.03	-0.08	0.03	0.01
	CB			0.76	-0.02	0.00	-0.02	0.02	0.00	0.01	0.01
	N	1.25	0.13	1.37	-0.06	-0.06	-0.07	0.00	0.06	-0.05	0.00
	O	0.66	0.08	0.73	0.00	0.00	-0.02	0.00	0.01	-0.02	-0.01
ARG	C	2.50	0.01[2]	2.51	-0.03	-0.07	-0.10	-0.04	0.11	-0.06	-0.04
	CA			2.04	0.08	0.06	0.07	-0.01	-0.08	0.01	-0.01
	CB			1.20	0.01	0.01	-0.01	0.00	0.00	0.01	0.00
	CD			1.21	-0.02	0.01	0.01	0.01	-0.01	0.02	-0.02
	CG			1.18	0.02	0.01	-0.02	0.02	0.00	0.00	0.01
	CZ			2.30	-0.03	0.00	0.01	-0.01	0.00	-0.01	-0.01
	N	1.28	0.13	1.40	-0.06	-0.04	-0.06	0.06	0.03	-0.03	0.05
	NE	1.11	0.13	1.22	-0.02	0.00	0.01	-0.02	0.00	-0.03	0.01
	NH1	0.78	0.04	0.82	0.02	0.01	0.00	-0.01	0.00	-0.01	0.00
	NH2	0.76	0.04	0.80	0.03	0.00	-0.01	0.01	0.00	0.00	0.01
	O	0.66	0.08	0.74	-0.02	-0.01	-0.02	-0.01	0.03	-0.01	-0.02
ASN	C	2.46	0.04[2]	2.48	0.05	0.01	-0.04	-0.11	0.08	-0.03	-0.09
	CA			2.07	0.10	0.03	0.04	-0.02	-0.07	0.02	-0.02
	CB			1.19	0.01	0.01	-0.01	-0.01	-0.01	0.02	0.00
	CG			2.31	0.07	0.00	-0.02	-0.03	0.00	0.03	0.01
	N	1.31	0.11	1.40	-0.05	-0.05	-0.07	0.06	0.06	-0.03	0.03
	ND2	0.86	-0.06	0.81	-0.02	0.01	-0.01	0.01	-0.02	0.00	0.03
	O	0.67	0.08	0.74	0.01	0.01	-0.03	-0.02	0.01	0.00	0.00
	OD1	0.62	0.10	0.71	0.00	0.03	-0.01	-0.02	-0.05	0.02	0.01
ASP	C	2.47	0.05[2]	2.51	0.06	0.02	-0.05	-0.08	0.07	-0.05	-0.10
	CA			2.05	0.05	0.02	0.03	-0.01	-0.06	0.01	0.00
	CB			1.16	0.04	0.00	-0.01	-0.02	0.01	0.00	0.00
	CG			2.33	0.01	0.00	-0.03	0.00	0.02	0.02	-0.01
	N	1.31	0.09	1.39	-0.07	-0.06	-0.07	0.06	0.05	-0.01	0.02
	O	0.67	0.09	0.75	0.02	0.01	-0.02	-0.02	0.01	0.00	-0.02
	OD1	0.66	0.08	0.74	0.01	0.03	-0.02	-0.01	-0.04	0.00	-0.01
	OD2	0.62	0.09	0.71	0.00	0.01	-0.02	-0.01	-0.01	0.02	0.01
CYS	C	2.47	0.02[2]	2.48	0.02	0.00	-0.07	0.00	0.10	-0.06	-0.04
	CA			2.04	0.09	0.03	0.06	-0.01	-0.12	0.05	-0.03
	CB			1.17	0.00	0.01	-0.01	0.00	0.00	0.01	0.01
	N	1.27	0.13	1.38	-0.02	-0.04	-0.04	0.07	0.05	-0.01	0.02
	O	0.67	0.06	0.73	0.02	0.00	-0.02	-0.01	0.02	0.00	-0.01
	SG			0.79	0.03	0.03	0.00	0.03	-0.04	0.03	0.00
GLN	C	2.49	0.03[2]	2.52	-0.05	-0.05	-0.09	-0.09	0.08	-0.06	-0.03
	CA			2.06	0.11	0.03	0.06	-0.05	-0.06	0.05	0.00
	CB			1.20	-0.02	0.01	-0.01	0.01	0.00	0.03	-0.01
	CD			2.28	0.06	0.01	0.00	0.01	-0.01	0.02	-0.01
	CG			1.18	-0.03	0.01	-0.01	-0.01	0.01	0.00	0.01
	N	1.33	0.09	1.41	-0.05	-0.04	-0.07	0.06	0.03	-0.02	0.06
	NE2	0.83	-0.05	0.79	0.01	0.00	-0.01	0.05	0.00	0.01	-0.01
	O	0.66	0.09	0.75	0.00	-0.02	-0.03	-0.03	0.03	-0.02	-0.04
	OE1	0.61	0.09	0.67	0.00	0.02	-0.01	0.02	-0.01	0.02	0.01

Table 19: Continued

Residue		No H Bonds Packing	1 H Bond ΔPack	Average Packing	Beta Bridge ΔPack	Random Coil[1] ΔPack	Beta Ladder ΔPack	3/10 Helix ΔPack	Alpha Helix ΔPack	Bend ΔPack	Turn ΔPack
GLU	C	2.51	0.03[2]	2.53	-0.05	-0.06	-0.11	-0.04	0.08	-0.06	0.00
	CA			2.05	0.09	0.06	0.06	-0.04	-0.05	0.05	-0.02
	CB			1.20	0.02	0.02	-0.01	0.01	0.00	0.01	0.00
	CD			2.31	0.10	0.05	0.00	0.00	-0.02	0.01	0.00
	CG			1.17	-0.01	0.02	-0.01	0.01	0.00	0.02	0.02
	N	1.36	0.06	1.41	-0.07	-0.04	-0.07	0.05	0.03	-0.03	0.03
	O	0.66	0.09	0.74	0.00	0.00	-0.03	-0.02	0.02	-0.02	-0.02
	OE1	0.63	0.07	0.70	-0.01	0.01	0.00	0.01	0.00	-0.01	0.00
	OE2	0.63	0.07	0.69	0.13	0.01	-0.01	0.00	0.00	0.02	-0.03
GLY	C	2.23	0.04[2]	2.26	0.04	-0.05	-0.06	0.07	0.15	-0.05	-0.03
	CA			1.18	0.03	0.02	0.03	-0.02	-0.05	0.00	0.00
	N	1.17	0.13	1.28	-0.02	-0.02	-0.05	0.03	0.06	0.00	0.01
	O	0.65	0.07	0.72	0.00	0.00	-0.01	0.00	0.02	0.00	-0.02
HIS	C	2.48	0.03[2]	2.51	-0.02	-0.05	-0.08	-0.06	0.12	-0.05	-0.08
	CA			2.03	0.06	0.03	0.05	-0.02	-0.08	0.02	-0.02
	CB			1.17	0.01	0.01	0.00	-0.02	-0.01	0.01	0.02
	CD2			1.11	0.04	0.03	-0.02	-0.01	-0.02	0.03	0.02
	CE1			1.13	0.04	0.00	0.00	0.06	0.00	-0.02	0.00
	CG			2.24	-0.03	0.01	-0.02	0.00	0.00	0.02	0.01
	N	1.28	0.14	1.40	-0.05	-0.03	-0.06	0.06	0.04	-0.02	0.05
	ND1	0.85	0.16	0.98	0.04	0.04	-0.02	-0.02	-0.02	0.03	-0.02
	NE2	1.03	0.14	1.15	-0.06	0.02	0.00	-0.03	-0.02	0.01	0.01
	O	0.66	0.09	0.74	-0.01	0.01	-0.03	-0.02	0.03	-0.01	0.00
ILE	C	2.57	0.04[2]	2.60	-0.05	-0.06	-0.11	0.00	0.13	-0.06	-0.01
	CA			2.10	0.08	0.04	0.08	-0.05	-0.11	0.04	-0.02
	CB			1.96	0.03	0.02	-0.01	0.00	0.00	0.02	0.03
	CD1			0.74	0.00	0.01	0.00	0.01	0.00	0.01	0.02
	CG1			1.16	0.01	0.02	-0.02	0.02	0.01	0.04	0.03
	CG2			0.78	0.02	0.02	-0.01	0.02	0.00	0.02	0.00
	N	1.30	0.11	1.41	-0.03	-0.03	-0.03	0.02	0.04	-0.02	0.02
	O	0.65	0.09	0.74	-0.01	-0.02	-0.03	0.00	0.04	-0.01	-0.01
LEU	C	2.47	0.04[2]	2.50	-0.03	-0.07	-0.09	-0.05	0.09	-0.08	-0.08
	CA			2.08	0.08	0.06	0.06	-0.02	-0.06	0.05	0.00
	CB			1.21	0.01	0.01	-0.01	0.00	0.00	0.02	0.01
	CD1			0.75	0.01	0.01	0.00	0.00	-0.01	0.00	0.00
	CD2			0.75	0.01	0.01	0.00	0.00	-0.01	0.01	0.00
	CG			1.91	-0.01	0.01	-0.04	-0.01	0.02	0.01	0.02
	N	1.27	0.15	1.40	-0.07	-0.04	-0.05	0.04	0.04	-0.03	0.04
	O	0.64	0.10	0.74	0.00	-0.01	-0.03	-0.01	0.02	-0.02	-0.01
LYS	C	2.50	0.02[2]	2.52	-0.03	-0.07	-0.10	-0.04	0.10	-0.07	-0.01
	CA			2.05	0.10	0.06	0.06	-0.02	-0.08	0.02	0.00
	CB			1.19	0.01	0.01	-0.01	0.00	0.00	0.00	0.01
	CD			1.15	0.06	0.04	0.00	0.02	-0.02	0.01	-0.03
	CE			1.17	0.04	0.02	0.00	0.03	-0.01	-0.02	-0.01
	CG			1.18	0.02	0.01	-0.02	0.02	0.00	0.02	0.01
	N	1.30	0.12	1.40	-0.03	-0.03	-0.06	0.05	0.02	-0.02	0.03
	NZ	0.66	0.04	0.71	-0.03	0.00	0.03	-0.08	0.00	-0.04	-0.02
	O	0.65	0.09	0.74	-0.01	-0.01	-0.02	0.00	0.03	-0.01	-0.02

Table 19: Continued

Residue		No H Bonds Packing	1 H Bond ΔPack	Average Buried Packing	Beta Bridge ΔPack	Random Coil[1] ΔPack	Beta Ladder ΔPack	3/10 Helix ΔPack	Alpha Helix ΔPack	Bend ΔPack	Turn ΔPack
MET	C	2.48	0.04[2]	2.51	-0.06	-0.06	-0.10	-0.04	0.09	-0.07	-0.06
	CA			2.07	0.03	0.04	0.07	0.01	-0.06	0.03	0.02
	CB			1.19	0.00	0.01	-0.01	-0.01	0.00	0.03	0.01
	CE			0.77	0.01	0.02	-0.01	0.02	0.00	0.01	0.01
	CG			1.18	-0.05	0.01	-0.01	-0.01	0.00	0.02	0.00
	N	1.27	0.15	1.41	-0.05	-0.05	-0.07	0.04	0.05	-0.03	0.02
	O	0.65	0.09	0.74	0.00	-0.01	-0.03	-0.01	0.02	-0.03	-0.02
	SD	0.86	0.04	0.86	0.02	0.02	0.00	0.00	-0.01	0.01	0.00
PHE	C	2.48	0.05[2]	2.53	-0.06	-0.07	-0.12	-0.05	0.15	-0.11	-0.05
	CA			2.04	0.09	0.04	0.06	-0.01	-0.08	0.02	-0.01
	CB			1.17	0.00	0.00	-0.01	-0.01	0.01	0.01	0.00
	CD1			1.14	0.02	0.02	0.00	0.00	-0.01	0.04	0.01
	CD2			1.12	0.01	0.03	-0.01	0.01	-0.01	0.03	0.02
	CE1			1.08	0.01	0.01	0.00	-0.01	0.00	0.00	0.01
	CE2			1.07	0.01	0.01	0.00	0.00	0.00	0.01	-0.01
	CG			2.27	-0.02	0.01	-0.02	0.00	0.01	0.02	0.00
	CZ			1.08	0.01	0.00	0.00	-0.01	0.00	0.00	-0.01
	N	1.30	0.12	1.41	-0.06	-0.05	-0.04	0.07	0.04	0.00	0.05
	O	0.66	0.08	0.73	0.00	-0.02	-0.02	-0.01	0.03	-0.03	0.00
PRO	C	2.47	0.02[2]	2.49	0.05	-0.01	-0.04	0.03	0.05	-0.02	0.01
	CA			1.97	0.01	0.03	0.01	-0.06	-0.10	0.06	0.00
	CB			1.11	-0.04	0.01	-0.02	0.00	-0.02	0.01	0.00
	CD			1.19	-0.07	-0.01	-0.03	0.04	0.07	-0.04	0.00
	CG			1.09	0.00	0.00	0.00	-0.03	-0.01	0.02	0.01
	N			2.16	-0.02	-0.02	-0.01	0.05	0.01	0.00	0.01
	O	0.66	0.09	0.73	0.02	0.00	-0.02	-0.01	0.02	0.00	-0.03
SER	C	2.44	0.03[2]	2.46	0.01	0.01	-0.09	-0.02	0.09	-0.02	-0.03
	CA			2.01	0.06	0.02	0.07	-0.04	-0.10	-0.01	-0.03
	CB			1.15	-0.02	0.01	-0.02	-0.01	0.01	0.01	0.00
	N	1.26	0.10	1.35	-0.04	-0.04	-0.04	0.04	0.06	-0.02	0.03
	O	0.66	0.08	0.74	0.02	0.02	-0.02	-0.02	0.01	-0.01	-0.01
	OG	0.55	0.10	0.65	0.02	0.01	-0.01	-0.01	-0.01	0.01	0.01
THR	C	2.51	0.03[2]	2.53	-0.04	-0.05	-0.08	0.02	0.13	-0.04	-0.05
	CA			2.07	0.08	0.00	0.07	-0.07	-0.12	0.00	-0.04
	CB			1.91	0.02	0.02	-0.01	0.01	-0.01	0.02	0.01
	CG2			0.77	0.01	0.02	-0.01	0.01	-0.01	0.01	0.00
	N	1.31	0.09	1.38	-0.02	-0.02	-0.05	0.05	0.04	-0.01	0.04
	O	0.67	0.08	0.74	0.01	0.02	-0.03	-0.04	0.02	0.00	-0.02
	OG1	0.59	0.09	0.67	0.03	0.01	0.00	-0.01	-0.02	0.01	0.01
TRP	C	2.51	0.04[2]	2.54	-0.05	-0.08	-0.11	-0.04	0.13	-0.07	-0.03
	CA			2.04	0.06	0.01	0.06	-0.05	-0.05	-0.02	-0.01
	CB			1.16	-0.01	0.00	-0.01	0.03	0.00	0.03	0.01
	CD1			1.11	0.06	0.03	0.00	0.02	-0.02	0.03	-0.01
	CD2			2.16	-0.01	0.02	-0.03	0.02	0.00	0.03	0.05
	CE2			2.21	-0.02	0.03	-0.01	0.03	-0.01	0.00	0.01
	CE3			1.12	0.05	0.02	0.02	0.01	-0.03	0.03	0.00
	CG			2.21	-0.02	0.00	-0.03	0.01	0.01	-0.01	0.07
	CH2			1.09	0.00	0.02	0.00	0.01	-0.01	0.00	0.00
	CZ2			1.10	0.01	0.01	0.01	0.01	-0.01	0.01	-0.02
	CZ3			1.07	0.00	0.03	0.01	-0.01	-0.02	0.02	0.01
	N	1.31	0.10	1.40	-0.07	-0.04	-0.05	0.07	0.04	-0.02	0.03
	NE1	1.02	0.11	1.11	0.00	0.02	0.00	-0.01	0.00	-0.03	-0.03
	O	0.66	0.09	0.75	-0.03	-0.02	-0.03	-0.01	0.03	-0.01	-0.02

Table 19: Continued

Residue		No H Bonds Packing	1 H Bond ΔPack	Average Buried Packing	Beta Bridge ΔPack	Random Coil[1] ΔPack	Beta Ladder ΔPack	3/10 Helix ΔPack	Alpha Helix ΔPack	Bend ΔPack	Turn ΔPack
TYR	C[2]	2.50	0.03[2]	2.52	-0.05	-0.04	-0.11	-0.05	0.15	-0.09	-0.04
	CA			2.05	0.07	0.03	0.05	-0.04	-0.09	0.02	-0.01
	CB			1.17	0.00	0.02	-0.01	0.02	0.00	0.02	0.01
	CD1			1.14	0.02	0.01	0.00	-0.01	-0.02	0.02	0.02
	CD2			1.13	0.04	0.02	-0.01	0.01	-0.01	0.03	0.02
	CE1			1.11	-0.03	0.01	0.00	0.00	-0.01	0.02	0.02
	CE2			1.11	0.02	0.02	0.00	0.02	-0.02	0.02	0.00
	CG			2.26	0.00	0.01	-0.01	-0.02	0.01	0.03	0.00
	CZ			2.22	0.00	0.02	0.00	-0.01	-0.01	0.01	0.01
	N	1.29	0.13	1.40	-0.03	-0.05	-0.03	0.08	0.03	0.00	0.05
	O	0.66	0.08	0.74	0.00	-0.01	-0.03	-0.01	0.03	0.01	-0.01
	OH	0.54	0.11	0.64	0.00	0.01	0.00	0.00	-0.01	0.03	0.01
VAL	C	2.55	0.03[2]	2.58	-0.04	-0.04	-0.10	0.00	0.14	-0.05	0.01
	CA			2.09	0.06	0.02	0.07	-0.06	-0.12	0.02	-0.03
	CB			1.90	0.04	0.02	-0.01	0.00	-0.01	0.02	0.03
	CG1			0.77	0.01	0.02	-0.01	0.00	0.00	0.02	0.01
	CG2			0.77	0.00	0.02	-0.01	0.04	0.00	0.03	0.03
	N	1.30	0.10	1.40	-0.03	-0.03	-0.02	0.03	0.04	-0.02	0.02
	O	0.65	0.08	0.73	0.00	-0.02	-0.03	0.00	0.04	0.00	-0.02

$\Delta Pack = Packing_{No\ Hbonds} - Packing_{Hbonds}$ or $\Delta Pack = Packing_{other} - Packing_{average}$

[1] No secondary structure was assigned by DSSP.

[2] The effect of the hydrogen bonding of the carbonyl oxygen at the carbonyl carbon

Discussion

Many previous studies have calculated the volume of amino acids in folded proteins (Richards 1974; Chothia 1975; Harpaz et al. 1994; Pontius et al. 1996; Tsai and Gerstein 2002). The volumes calculated in this study show good agreement with the total volumes for residues determined in previous studies as shown in Table 20. Differences in the data sets and radii used in the calculations contribute to the minor differences in the results of different groups. The results from Pontius et al (Pontius et al. 1996) are generally larger than others since they were determined using a classical Voronoi instead of the modified Voronoi which takes into account differences in the radii between atom groups.

Table 20: Comparison of residue volumes to other studies

	Backbone	Side Chain	Total	Richards[1]	Chothia[2]	Harpaz[3]	Pontius[4]	Tsai[5]
ALA	53.0	36.2	89.2	90.6	90.1	90.1	91.5	89.3
ARG	51.7	141.6	193.3	198.0	193.5	192.8	196.1	190.3
ASN	52.8	74.8	127.6	126.0	135.2	127.5	138.3	122.4
ASP	52.3	66.4	118.6	118.1	116.9	117.1	135.2	114.4
CYS	52.6	58.9	111.6	115.3	113.9	113.2	114.4	112.8
GLN	51.9	98.9	150.8	149.5	148.5	149.4	156.4	146.9
GLU	52.3	90.3	142.6	142.0	140.9	140.8	154.6	138.8
GLY	64.3		64.3	64.1	64.4	63.8	67.5	63.8
HIS	52.6	107.9	160.5	157.9	158.4	159.3	163.2	157.5
ILE	51.4	111.0	162.4	165.7	163.6	164.9	162.6	163.0
LEU	52.0	111.3	163.3	165.7	164.0	164.6	163.4	163.1
LYS	52.0	114.5	166.5	178.3	168.0	170.0	162.5	165.1
MET	51.9	114.5	166.4	166.5	166.8	167.7	165.9	165.8
PHE	52.2	139.8	192.0	192.4	192.5	193.5	198.8	190.8
PRO	48.0	73.7	121.6	123.8	122.7	123.1	123.4	121.6
SER	53.1	42.1	95.2	97.2	94.4	94.2	102.0	94.2
THR	52.0	68.0	120.0	122.6	119.8	120.0	126.0	119.6
TRP	52.3	179.0	231.3	228.5	229.8	231.7	237.2	226.4
TYR	52.2	145.5	197.7	197.9	195.9	197.1	209.8	194.6
VAL	51.6	86.4	138.0	140.4	139.0	139.1	138.4	138.2

All values in $Å^3$; Values reported from previous studies are the total residue volume.
[1] (Richards 1974)
[2] (Chothia 1975)
[3] (Harpaz et al. 1994)
[4] (Pontius et al. 1996)
[5] (Tsai and Gerstein 2002)

Hydrogen Bonding

Hydrogen bonding increases the local packing in the protein interior. This effect is seen in both the volume of hydrogen bonded amino acids and the packing of hydrogen bonded atom groups. For amino acids with two hydrogen bonds, the volume change (3.5 $Å^3$) is slightly more than twice the volume change for one hydrogen bond (1.4 $Å^3$).

Also, there is a small but consistent increase (0.03) in the packing of backbone carbonyl carbons when there is hydrogen bonding by the carbonyl oxygen. It appears that the increased local packing also affects atoms in vicinity to the hydrogen bond. By increasing the local packing around polar groups, there is an increase in the van der Waals interactions for those atoms. It is possible that the increased van der Waals interactions of buried polar groups can compensate for the solvation penalty for removing those polar groups from the aqueous environment. The increased van der Waals interactions is in addition to the favorable electrostatic interactions of the hydrogen bond.

Conclusion

Using Voronoi polyhedra, I have calculated the volume occupied by individual atoms groups in 872 high resolution crystal structures. These data have been analyzed to investigate the effect of hydrogen bonding and secondary structure on local packing. Hydrogen bonding increases local packing, reducing backbone volumes by approximately 1.5 $Å^3$ per hydrogen bond.

STRUCTURAL COMPARISON OF RNASE SA AND 5K VARIANT

Introduction

Electrostatic interactions can play an important role in protein stability. Altering the pH of the solution has long been known to denature proteins (Anson and Mirsky 1931; Anson 1945; Lumry and Eyring 1954). Changes in salt concentration have also been shown to effect protein stability (Acampora and Hermans 1967). Both of these methods alter the electrostatic environment of the protein, by titrating charged groups or screening electrostatic interactions amoung them.

In 2001, Shaw et al reported a variant of RNase Sa where the five surface acidic groups (aspartic acid and glutamic acid) shown in Figure 15 were replaced with five basic lysine groups (Shaw et al. 2001). The wild type RNase Sa has a pI (the pH where the net charge of the protein is zero) of 3.5, while the 5K variant (D1K+D17K+D25K+E41K+E74K) has a pI above 10. Even with the drastic change in charge, there is little effect on the stability of the protein, with both the magnitude and pH dependence being nearly identical. Comparisons of ^{1}H and ^{15}N chemical shifts between the two proteins show relatively small changes (Laurents et al. 2003), indicative of only minor perturbations in the structure. Recently the X-ray structure of the 5K variant has been determined (Takano et al, unpublished data). Here, I present a detailed comparison of the 5K structure with previously determined structures of RNase Sa.

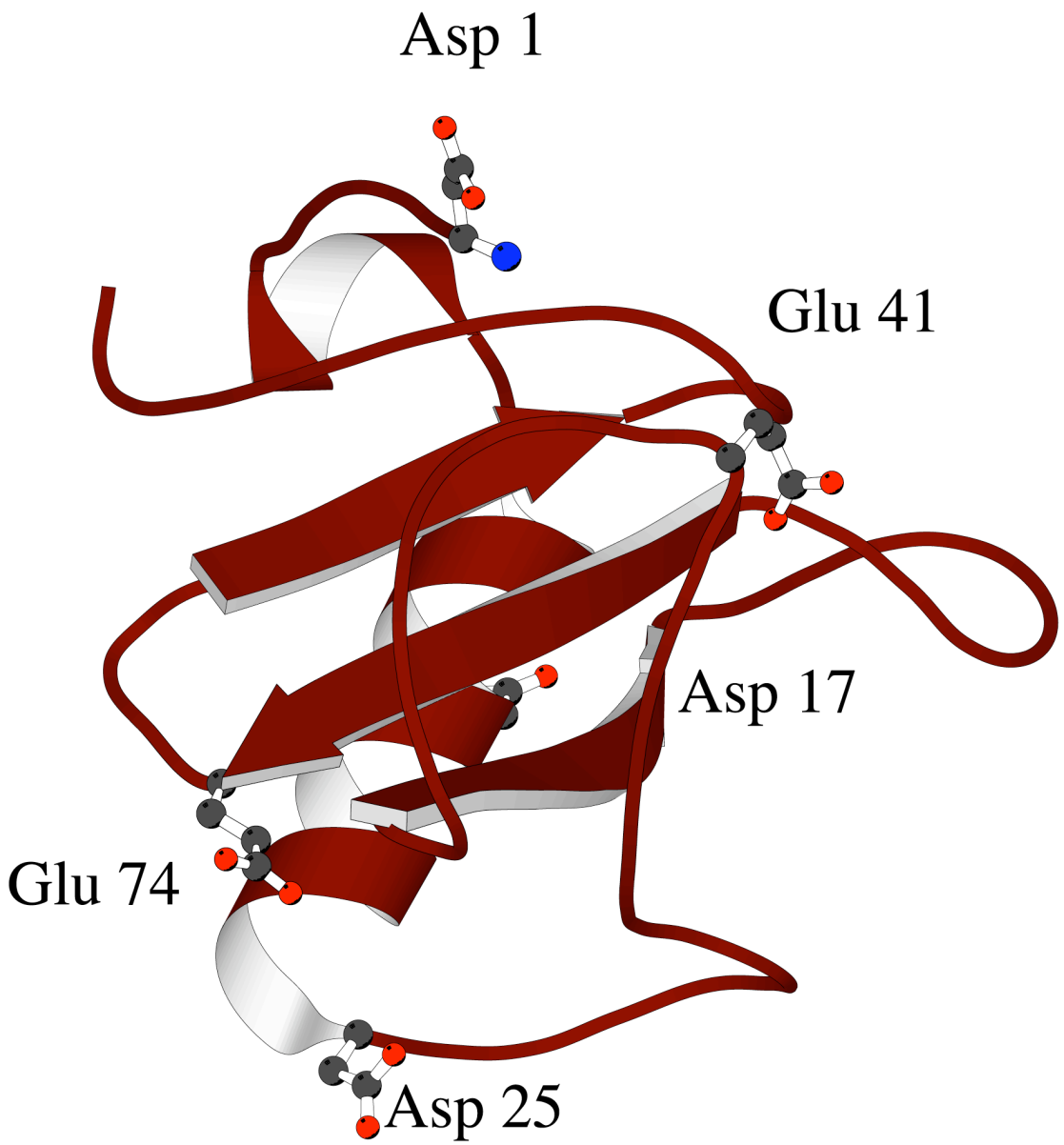


Figure 15: Location of lysine substitutions in the 5K variant. The figure was produced with Molscript (Kraulis 1991).

In addition, I use trans hydrogen bond J-coupling to compare the backbone hydrogen bonds of RNase Sa and the 5K variant in solution. Trans H-bond J-coupling is the only known method capable of experimentally linking the donor atom with the acceptor atom. These couplings were first observed in proteins between backbone amide protons and coordinated metal ions (Blake et al. 1992). In 1998, Dingley and Grzesiek demonstrated J-couplings between the acceptor and donor atoms across H-bonded base pairs in double stranded DNA (Dingley and Grzesiek 1998). More recently these couplings have been detected in hydrogen bonds in proteins (Cordier and Grzesiek 1999; Cornilescu et al. 1999a). In 1999, Cornilescu et al demonstrated a strong correlation between the magnitude of the J-coupling and the distance between the donor and acceptor atoms measured from crystal structures (Cornilescu et al. 1999b).

14 $$R_{NO} = 2.75 - 0.25\ln(-{}^{3h}J_{NC'}) \pm 0.06\text{Å}$$

In 2000, Cordier et al used this empirical relationship to observe changes in the donor-acceptor length of backbone-backbone hydrogen bonds in c-Src SH3 domain upon binding to the high affinity ligand RLP2 (Cordier et al. 2000). They demonstrated that trans H-bond J-couplings could be used to accurately measure changes in H-bond lengths smaller then 0.12 Å.

The reversal of the net charge in the RNase 5K variant provides a unique opportunity to investigate the electrostatic nature of protein stability. Specific electrostatic interactions have not been altered, but the total electrostatic environment of the protein has

completely changed. The result is a system that can be used to investigate the effect electrostatics on all aspects of protein stability.

Methods

Protein Expression and Purification

RNase Sa and the 5K variant were expressed as described by Laurents et al (Laurents et al. 1999) with some modification. *Escherichia coli* (strain RY1988) harboring the expression vector pEH100 was grown in M9 minimal media with increasing concentrations of D_2O. All cultures were grown at 37°C and with 25 μg/ml of ampicillin unless otherwise noted. Initially, a 30ml culture in 20% D_2O was grown for 24 hours. Another 30 ml culture containing 50% D_2O was inoculated from the 20% D_2O culture and was grown an addition 24 hours. A final 30 ml culture containing 80% D_2O was inoculated from the 50% D_2O culture and grown for 24 hours. The 80% D_2O culture was used to inoculate twelve 2L flasks each containing 500 ml of M9 minimal media containing 98% D_2O. The cells were grown to an OD of 1.2-1.5. At this time, the media was replaced with 3L of M9 minimal media containing ^{15}N-NH_4Cl and ^{13}C-glucose and 98% D_2O and the culture was induced with 0.1 mM IPTG and the cells were incubated in six 2L flasks at 30°C for 12 hours. RNase Sa was purified as described by Hebert et al (Hebert et al. 1997) and the 5K variant was purified as described by Shaw et al (Shaw et al. 2001).

NMR

All data were collected on a 600 Mhz Varian Inova with triple axis gradients at 30°C. Samples were prepared in 50 mM Phosphate buffer, pH 5.5, with 10% D_2O. The samples contained 0.1% sodium azide to prevent microbial growth and 1mM DSS as an internal chemical shift reference. The spectra were analyzed using nmrPipe (Delaglio et al. 1995) and PIPP (Garrett et al. 1991). To detect the trans hydrogen bond J-coupling, a long-range, water flip-back HNCO experiment with a transfer time of 2T=133 ms was used as described by Cordier and Grzesiek (Cordier and Grzesiek 1999). The sensitivity was enhanced by using composite-pulse decoupling as described by Liu et al (Liu et al. 2000).

Computational Analysis

Using the X-ray crystal structures, HBPLUS (McDonald and Thornton 1994) and DSSP (Kabsch and Sander 1983) were used to determine the hydrogen bonding and the secondary structure assignments. The RMSDs were determined using the code-mbg library (Gerstein 1992).

Results and Discussion

There are two high-resolution structures available for RNase Sa. 1rgg was determined at room temperature to a resolution of 1.2 Å while 1lni was determined at 100 K to a resolution of 1.0 Å. Both structures were determined from crystals grown under similar

solvent conditions of phosphate buffer with ammonium sulfate precipitation, and have identical crystal forms with two molecules in the unit cell. The 5K structure was determined to a resolution of 1.8 Å at room temperature, but with different solvent conditions, acetate and Tris buffer, and contains only one molecule per unit cell. Table 21 shows the differences in the crystal structure parameters. With the two molecules per unit cell for each of the RNase Sa structures, there are four molecules to compare to the 5K structure. Since molecule A of 1rgg has the smallest Cα RMSD to the 5K structure, it will be used as the basis for comparison.

Table 21: Comparison of crystal structures

	1rgg	1lni	5K
Resolution	1.2 Å	1.0 Å	1.8 Å
Space Group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Temperature	r.t.	100 K	r.t.
Unit Cell			
Dimension (a,b,c)	64.7, 78.6, 39.0	64.2, 77.8, 38.3	51.3, 36.7, 46.2
Angles (α,β,γ)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Molecules	2	2	1

Secondary Structure

The secondary structure, shown in Table 22, between RNase Sa and the 5K variant is nearly identical, but there are a few differences. The D1K mutation in 5K seems to have improved the pairing of the N-terminal and C-terminal beta strands. The N-terminal beta strand in the 5K structures extends from residue 2 to residue 7, whereas in the RNase Sa structure the strand does not begin until residue 4. The C-terminal beta strand in 5K is

extended one residue, starting at residue 89 instead of residue 90 in RNase Sa. The D25K mutation in 5K appears to have had an effect on the C-terminus of the major helix, resulting in a shortening of the helix by one residue. The shortening of the major helix by one residue is also seen in molecule A of 1lni.

Table 22: Comparison of secondary structure

	1									10										20										30
Sequence	**D**	V	S	G	T	V	C	L	S	A	L	P	P	E	A	T	**D**	T	L	N	L	I	A	S	**D**	G	P	F	P	Y
1rgg Molecule A	**C**	C	C	C	E	E	E	G	G	G	S	C	H	H	H	H	**H**	H	H	H	H	H	H	H	**T**	C	C	C	S	S
5K	**C**	E	E	E	E	E	E	G	G	G	S	C	H	H	H	H	**H**	H	H	H	H	H	H	T	**T**	C	C	C	S	S
										40										50										60
Sequence	S	Q	D	G	V	V	F	Q	N	R	**E**	S	V	L	P	T	Q	S	Y	G	Y	Y	H	E	Y	T	V	I	T	P
1rgg Molecule A	T	T	T	T	C	B	C	C	C	T	**T**	C	C	S	C	C	C	C	T	T	S	C	E	E	E	E	C	C	C	T
5K	T	T	T	T	C	B	C	C	C	T	**T**	C	C	S	C	C	C	C	T	T	S	C	E	E	E	E	C	C	C	T
										70										80										90
Sequence	G	A	R	T	R	G	T	R	R	I	I	T	G	**E**	A	T	Q	E	D	Y	Y	T	G	D	H	Y	A	T	F	S
1rgg Molecule A	T	C	S	S	C	C	S	C	E	E	E	E	C	**S**	S	T	T	C	E	E	E	E	S	S	T	T	S	C	C	E
5K	T	C	S	S	C	C	S	C	E	E	E	E	C	**S**	S	T	T	C	E	E	E	E	S	S	T	T	S	S	E	E
										96																				
Sequence	L	I	D	Q	T	C																								
1rgg Molecule A	E	E	E	T	T	C																								
5K	E	E	E	T	T	C																								

Residues substituted with lysine in 5K are bold. B is an isolated beta bridge, C is a random coil where DSSP did not assign a secondary structure, E is an extended beta strand, G is a 3/10 helix, H is a alpha helix, S is a bend and T is a hydrogen bonded turn.

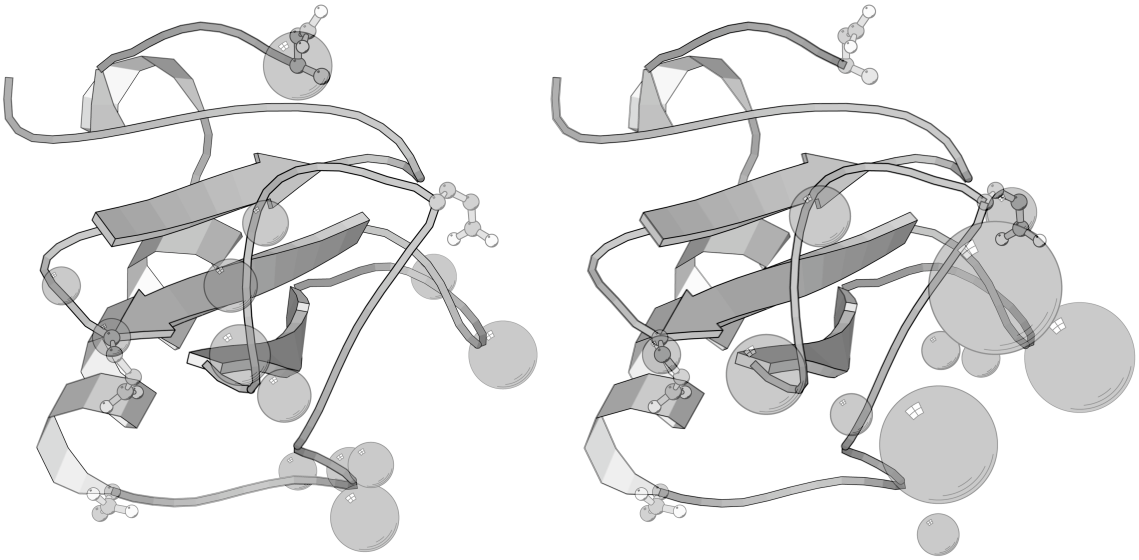


Figure 16: Differences in the crystal structures. A) Variation in the backbone positions. Amino acids with a Cα RMSD >0.5 Å are represented by spheres centered at the Cα and proportional to the RMSD. B) Variation in the side chain positions. Amino acids with an average side chain RMSD>1.0 Å are represented by spheres centered at the Cβ and proportional to the average RMSD. The figure was produced with MolScript (Kraulis 1991).

Table 23: RMSD between RNase Sa and the 5K variant

Residue		Ca RMSD	All Atom RMSD			Residue		Ca RMSD	All Atom RMSD		
			Backbone	Sidechain	Amino Acid				Backbone	Side chain	Amino Acid
1	**ASP**	**0.98**	**1.06**	**0.67**	**0.93**	50	GLY	0.47	0.38		0.38
2	VAL	0.29	0.30	0.76	0.50	51	TYR	0.24	0.27	0.25	0.26
3	SER	0.44	1.00	0.60	0.86	52	TYR	0.22	0.22	0.30	0.27
4	GLY	0.39	0.40		0.40	53	HIS	0.22	0.27	0.30	0.29
5	THR	0.31	0.32	0.35	0.33	54	GLU	0.31	0.34	0.93	0.54
6	VAL	0.30	0.31	0.45	0.37	55	TYR	0.13	0.20	0.15	0.17
7	CYS	0.18	0.18	0.23	0.20	56	THR	0.09	0.09	0.12	0.10
8	LEU	0.17	0.15	0.20	0.18	57	VAL	0.05	0.16	0.10	0.14
9	SER	0.12	0.12	0.25	0.16	58	ILE	0.13	0.12	0.15	0.13
10	ALA	0.26	0.26	0.42	0.30	59	THR	0.14	0.11	0.08	0.10
11	LEU	0.28	0.22	0.26	0.24	60	PRO	0.24	0.32	0.32	0.32
12	PRO	0.26	0.22	0.27	0.24	61	GLY	0.66	0.44		0.44
13	PRO	0.31	0.28	0.31	0.29	62	ALA	0.36	0.46	0.48	0.47
14	GLU	0.19	0.12	0.19	0.15	63	ARG	0.91	1.05	2.96	2.26
15	ALA	0.14	0.13	0.16	0.14	64	THR	0.46	0.55	0.67	0.60
16	THR	0.12	0.09	0.12	0.10	65	ARG	0.42	0.41	1.00	0.79
17	**ASP**	**0.02**	**0.06**	**0.18**	**0.10**	66	GLY	0.34	0.34		0.34
18	THR	0.12	0.14	0.14	0.14	67	THR	0.04	0.20	0.23	0.21
19	LEU	0.13	0.18	0.18	0.18	68	ARG	0.13	0.19	0.49	0.38
20	ASN	0.17	0.23	0.54	0.39	69	ARG	0.07	0.12	0.34	0.26
21	LEU	0.18	0.25	0.31	0.28	70	ILE	0.11	0.21	0.15	0.18
22	ILE	0.26	0.32	0.29	0.30	71	ILE	0.13	0.16	0.37	0.26
23	ALA	0.29	0.30	0.39	0.32	72	THR	0.10	0.18	0.05	0.12
24	SER	0.16	0.25	0.12	0.21	73	GLY	0.39	0.30		0.30
25	**ASP**	**0.12**	**0.25**	**0.89**	**0.46**	**74**	**GLU**	**0.54**	**0.48**	**1.08**	**0.68**
26	GLY	0.18	0.24		0.24	75	ALA	0.50	0.41	0.58	0.44
27	PRO	0.17	0.30	0.31	0.31	76	THR	0.37	0.37	0.41	0.38
28	PHE	0.44	0.45	0.32	0.37	77	GLN	0.36	0.26	0.87	0.60
29	PRO	0.54	0.58	0.58	0.58	78	GLU	0.25	0.23	0.43	0.30
30	TYR	0.63	0.65	0.76	0.72	79	ASP	0.21	0.24	0.23	0.24
31	SER	0.98	0.95	1.19	1.03	80	TYR	0.23	0.24	0.31	0.29
32	GLN	0.62	0.76	3.10	2.06	81	TYR	0.19	0.20	0.20	0.20
33	ASP	0.18	0.30	0.17	0.25	82	THR	0.16	0.18	0.26	0.22
34	GLY	0.18	0.19		0.19	83	GLY	0.08	0.07		0.07
35	VAL	0.12	0.14	0.42	0.26	84	ASP	0.21	0.30	0.25	0.28
36	VAL	0.17	0.19	0.28	0.23	85	HIS	0.18	0.27	1.38	0.94
37	PHE	0.22	0.20	0.24	0.22	86	TYR	0.33	0.29	0.33	0.32
38	GLN	0.21	0.21	1.03	0.67	87	ALA	0.20	0.18	0.33	0.21
39	ASN	0.40	0.35	0.37	0.36	88	THR	0.20	0.21	0.20	0.21
40	ARG	0.40	0.40	3.59	2.43	89	PHE	0.29	0.28	0.50	0.42
41	**GLU**	**0.35**	**0.36**	**0.54**	**0.42**	90	SER	0.15	0.25	0.09	0.20
42	SER	0.39	0.33	0.40	0.35	91	LEU	0.32	0.31	0.32	0.32
43	VAL	0.28	0.28	0.44	0.35	92	ILE	0.31	0.31	0.29	0.30
44	LEU	0.23	0.30	0.20	0.25	93	ASP	0.15	0.21	0.15	0.19
45	PRO	0.31	0.26	0.35	0.30	94	GLN	0.15	0.12	0.11	0.11
46	THR	0.60	0.62	1.67	1.07	95	THR	0.11	0.16	0.14	0.15
47	GLN	0.72	0.76	0.69	0.72	96	CYS	0.26	0.27	0.31	0.28
48	SER	0.80	0.81	2.16	1.26	Average		0.36			0.77
49	TYR	0.75	0.68	1.10	0.96						

Residues substituted with lysine in 5K are bold. At the substituted positions, the all atom RMSD was calculated using Cβ and Cγ for aspartic acid and Cβ, Cγ, and Cδ for glutamic acid.

Crystal Structures

The positional error associated with a crystal structure is approximately 1/10 the resolution. Since 1rgg has a 1.2 Å resolution (0.12 Å error) and the 5K structure has a 1.8 Å resolution (0.18 Å error), the error between the structures should be less than 0.3 Å. To understand where the largest changes in the backbone conformation occurred, a CαRMSD was calculated between molecula A in 1rgg and the 5K structures. For a better understanding of how the mutations affected packing of the side chains, an all atom RMSD (aaRMSD) was calculated between the two structures. Table 23 shows the RMSDs for each amino acid and Figure 16 maps the regions of high RMSD to the 3-dimensional structure of RNase Sa. Neither the D17K mutation, on the back of the major helix, or the D25K mutation, at the C-terminus of the major alpha helix, has a significant effect on the local structure. The secondary structure assignment demonstrated a change in the N-terminal beta sheet. This is manifested in a 1.06 Å RMSD for the first Cα. The E74K mutation, at the C-terminus of the middle strand, and the E41K mutation, in the major loop, both appear to cause significant rearrangement of the neighboring loops. In the region around the E41K mutation, there is significant reorganization of the side chains. The aaRMSD for the involved side chains (residues 31, 32, 63, 64, and 65) ranges from 0.67 Å to 3.10 Å. The repacking of the side chains is associated with movement in the backbone with CαRMSDs ranging from 0.36 to 0.98 Å for residues 29 to 32 and 61 to 63. The E74K mutation causes some local backbone rearrangement for residues 74 and 75 (CαRMSDs of 0.5 and 0.54 Å). There is also a change in residues 46 through 49. The side chain aaRMSDs range from 0.69 to 2.16 Å and the backbone

CαRMSDs range from 0.60 to 0.80 Å. It should be noted that these loops form the interface with molecule B in the RNase Sa structures and these differences may be due to crystal packing differences between the RNase Sa and 5K structures.

Hydrogen Bonding

The hydrogen bonds have been divided into two groups. Those where the backbone amide acts as the donor and those where side chain atoms act as the donor. Table 24 is a comparison of the hydrogen bonds donated by the backbone amide. Out of 55 hydrogen bonds, there are nine where a hydrogen bond is observed in one structure but not in the other. Five of these are within one amino acid of a mutation site. Table 25 shows a comparison of the hydrogen bonds donated by side chain atoms. Out of 29 hydrogen bonds, there are 12 where a hydrogen bond is observed in one structure but not in the other. Many of the changes in the hydrogen bonding are localized to areas near the mutation sites, but a few changes are distant from mutations. The loop from residue 84 to 89 has several changes in the hydrogen bonding. Aspartic acid 84 loses a hydrogen bond to threonine 82, but gains hydrogen bonds to the backbone of alanine 87. This region also sees changes in the hydrogen bonding of the side chains of histidine 85, tyrosine 86 and serine 90. There are also changes in the hydrogen bonding in regions that are close in space to E42K. The side chain of arginine 65 loses a hydrogen bond while threonine 67 gains a hydrogen bond and there is a change in the hydrogen bonding of the backbone amides of glutamine 32 and valine 35.

Table 24: Backbone amide hydrogen bonds

Donor		RNase Sa Acceptor			5K Acceptor			Donor		RNase Sa Acceptor			5K Acceptor		
		Residue		Atom	Residue		Atom			Residue		Atom	Residue		Atom
1	**ASP**	43	VAL	O				53	HIS	72	THR	O	72	THR	O
3	SER	89	PHE	O	89	PHE	O	54	GLU	35	VAL	O	35	VAL	O
6	VAL	91	LEU	O	91	LEU	O	55	TYR	70	ILE	O	70	ILE	O
8	LEU	93	ASP	O	93	ASP	O	56	THR	33	ASP	Oδ1	33	ASP	Oδ1
9	SER	96	CYS	O	96	CYS	O	57	VAL	68	ARG	O	68	ARG	O
10	ALA	7	CYS	O	7	CYS	O	62	ALA	59	THR	O	59	THR	O
11	LEU	8	LEU	O	8	LEU	O	66	GLY	64	THR	Oγ1	64	THR	Oγ1
15	ALA	12	PRO	O	12	PRO	O	69	ARG	82	THR	O	82	THR	O
16	THR	13	PRO	O				70	ILE	55	TYR	O	55	TYR	O
17	**ASP**				13	PRO	O	71	ILE	80	TYR	O	80	TYR	O
18	THR	14	GLU	O	14	GLU	O	72	THR	53	HIS	O	53	HIS	O
19	LEU	15	ALA	O	15	ALA	O	73	GLY	78	GLU	O	78	GLU	O
20	ASN	16	THR	O	16	THR	O	75	ALA				78	GLU	Oε1
21	LEU	17	ASP	O	17	LYS	O	78	GLU	75	ALA	O	75	ALA	O
22	ILE	18	THR	O	18	THR	O	80	TYR	71	ILE	O	71	ILE	O
23	ALA	19	LEU	O	19	LEU	O	81	TYR	90	SER	O	90	SER	O
24	SER				20	ASN	O	82	THR	69	ARG	O	69	ARG	O
25	**ASP**	22	ILE	O	22	ILE	O	84	ASP	82	THR	Oγ1			
26	GLY	21	LEU	O	21	LEU	O	85	HIS	67	THR	O	67	THR	O
30	TYR	33	ASP	Oδ2	33	ASP	Oδ2	86	TYR	82	THR	Oγ1	82	THR	Oγ1
32	GLN	32	GLN	Oε1				87	ALA				84	ASP	O/Oδ1
33	ASP	30	TYR	O	30	TYR	O	88	THR	84	ASP	Oδ1	84	ASP	Oδ1
34	GLY	54	GLU	O	54	GLU	O	90	SER	81	TYR	O	81	TYR	O
35	VAL	32	GLN	O				91	LEU	4	GLY	O	4	GLY	O
37	PHE	52	TYR	O	52	TYR	O	92	ILE	79	ASP	O	79	ASP	O
42	SER	39	ASN	O	39	ASN	O	93	ASP	6	VAL	O	6	VAL	O
44	LEU	39	ASN	Oδ1	39	ASN	Oδ1	95	THR	93	ASP	Oδ1	93	ASP	Oδ1
51	TYR	48	SER	O	48	SER	O								

Residues substituted with lysine in 5K are bold.

Table 25: Side chain donor hydrogen bonds

Donor			RNase Sa Acceptor		5K Acceptor		Donor			RNase Sa Acceptor		5K Acceptor	
Residue		Atom	Residue	Atom	Reside	Atom	Residue		Atom	Residue	Atom	Reside	Atom
5	THR	Oγ1	91 LEU	O			67	THR	Oγ1			83 GLY	O
9	SER	Oγ			96 CYS	O	69	ARG	Nε	86 TYR	Oη	86 TYR	Oη
18	THR	Oγ1			14 GLU	O	69	ARG	Nη1	65 ARG	O	65 ARG	O
			56 THR	O	56 THR	O				68 ARG	O	68 ARG	O
24	SER	Oγ			20 ASN	O	69	ARG	Nη2	66 GLY	O	66 GLY	O
					21 LEU	O	80	TYR	Oη	78 GLU	Oε2	78 GLU	Oε2
			26 GLY	O	26 GLY	O	82	THR	Oγ1	84 ASP	Oδ1	84 ASP	Oδ1
39	ASN	Nδ2	44 LEU	O	44 LEU	O				88 THR	O	88 THR	O
40	ARG	Nη1	38 GLN	Oε1			85	HIS	Nδ1	85 HIS	O		
42	SER	Oγ			42 SER	O	86	TYR	Oη			54 GLU	Oε2
51	TYR	Oη	78 GLU	Oε1	78 GLU	Oε1	88	THR	Oγ1	84 ASP	Oδ2	84 ASP	Oδ2
52	TYR	Oη	45 PRO	O	45 PRO	O	90	SER	Oγ	3 SER	O		
56	THR	Oγ1	33 ASP	Oδ1	33 ASP	Oδ1	95	THR	Oγ1	93 ASP	Oδ1	93 ASP	Oδ1
65	ARG	Nε	56 THR	Oγ1	56 THR	Oγ1							
65	ARG	Nη2	54 GLU	Oε2									
			56 THR	Oγ1	56 THR	Oγ1							

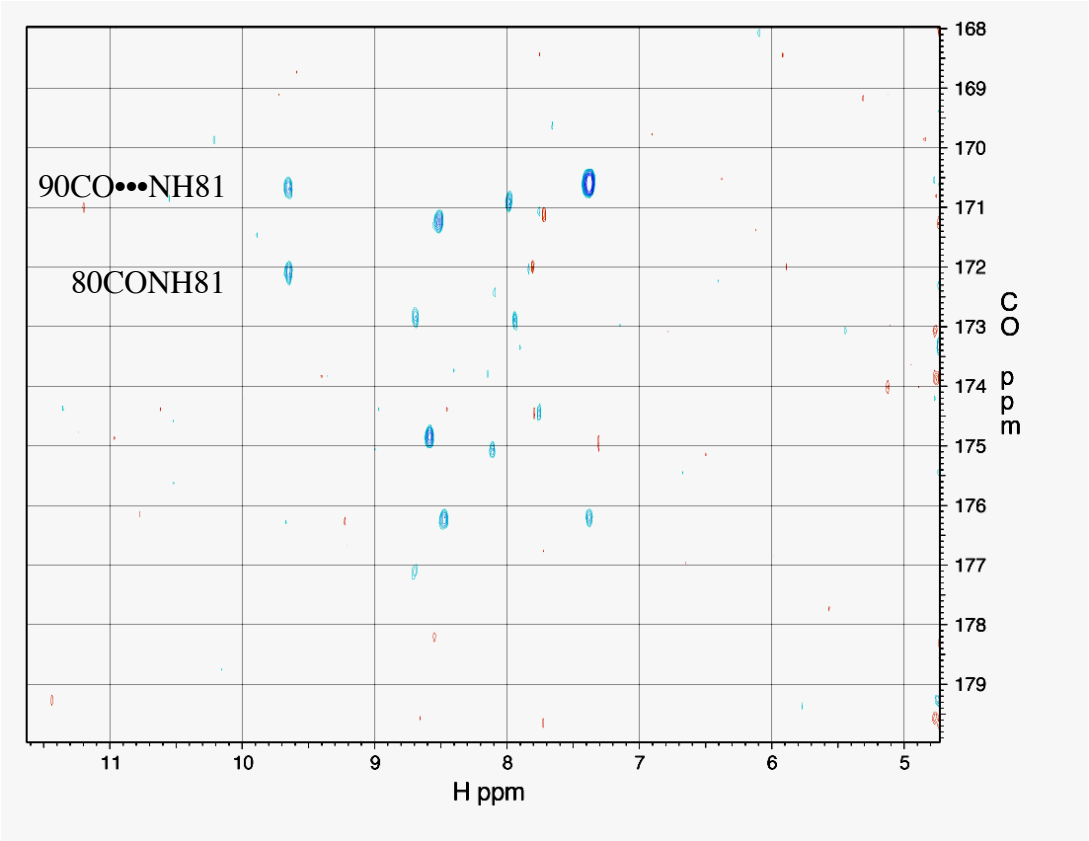


Figure 17: Sample long range HNCO used to determine trans hydrogen bond J couplings in RNase Sa.

In addition to measuring the hydrogen bonds in the X-ray crystal structures, NMR techniques were used to observed the hydrogen bonds in solution. While trans hydrogen bond J coupling can be used to accurately determine the donor-acceptor distance of the hydrogen bond, that proved to be difficult in this case. The J coupling between the amide nitrogen and the carbonyl carbon across the hydrogen bond is very week, approximately 0.5 Hz. The small J coupling results in a weak signal in the long-range HNCO experiment and thus a poor signal-to-noise ratio. The size of the protein and the concentration have an affect on the signal-to-noise ratio. Increasing the size of the protein results in an increase in line broadening, decreasing the overall peak height and thus the signal-to-noise. The signal-to-noise can be improved by increasing the concentration of the protein. Unfortunately, given the solution conditions, the solubility of the 5K variant achieved was approximately 0.5 mM. While other conditions could have been chosen to increase the solubility of the 5K variant, the solution conditions for both RNase Sa and the 5K variant should match for a meaningful comparison of the hydrogen bond lengths and changing the conditions to increase the solubility of the 5K variant would have reduced the solubility of wild-type RNase Sa. While the trans hydrogen bond J couplings could be observed, the error associated with the measured J couplings would have approximated the magnitude of the measured values, and any attempt to compare the donor-acceptor distances would have been meaningless.

Table 26 shows the observed trans hydrogen bond J-coupling in both RNase Sa and the 5K variant. Since RNase Sa is more soluble than the 5K variant under the conditions used for NMR, it is more difficult to observe hydrogen bonds in the 5K variant and some

of the hydrogen bonds observed in RNase Sa may form in the 5K variant but remain unobserved. Many of the hydrogen bonds observed by NMR are the same as would be expected from the crystal structures. Where there is a difference, the hydrogen bonds generally represent a single residue shift in the hydrogen-bonding pattern. For example, in the NMR analysis, a hydrogen bond is observed between the amide of glycine 26 and the carboxyl of isoleucine 22 in RNase Sa, but in the crystal structure, the carboxyl of isoleucine 22 is expected to hydrogen bond to aspartic acid 25 while the amide of glycine 26 forms a hydrogen bond with leucine 21. This hydrogen bond is not observed by NMR in the 5K variant, but the hydrogen bonding deduced from the crystal structures in this region shows no difference between RNase Sa and 5K. Another example is seen in 5K, where the carbonyl of tyrosine 52 is observed by NMR to form a hydrogen bond to glutamine 38, but not in RNase Sa. In both crystal structures, the carboxyl of tyrosine 52 is expected to form a hydrogen bond with phenylalanine 37. A unique hydrogen bond is observed from the amide of cysteine 96 to the carbonyl of leucine 8. In the crystal structure, a hydrogen bond is observed from the carbonyl of cysteine 96 to the amide of serine 9, so the hydrogen bond seen by NMR would be expected since these residues represent two hydrogen bonded strands of a beta sheet.

Table 26: Observed trans hydrogen bond J-coupling

Donor Residue		RNase Sa Acceptor Residue		5K Acceptor Residue	
Backbone-Backbone					
22	ILE	18	THR	18	THR
26	**GLY**	**22**	**ILE**		
34	GLY	54	GLU		
38	**GLN**			**52**	**TYR**
55	TYR	70	ILE		
70	ILE	55	TYR	55	TYR
80	TYR	71	ILE	71	ILE
81	TYR	90	SER	90	SER
90	SER	81	TYR		
96	**CYS**	**8**	**LEU**	**8**	**LEU**
Backbone-Side Chain					
17	**ASP**	**14**	**GLU**	**14**	**GLU**
24	**SER**	**20**	**ASN**	**20**	**ASN**
75	ALA			78	GLU
88	THR	84	ASP	84	ASP
93	**ASP**	**94**	**GLN**	**94**	**GLN**

Backbone-Backbone hydrogen bonds are between the donor amine and the acceptor carbonyl. Backbone-Side chain hydrogen bonds are between the donor amide and the acceptor side chain oxygen. Hydrogen bonds in bold are not seen in the crystal structures.

Three backbone to side-chain hydrogen bonds are observed by NMR that are not expected from the crystal structures; aspartic acid 17 (lysine in 5K), serine 24, and aspartic acid 93. At the C-terminus of the major helix, the amide of serine 24 is observed to form a hydrogen bond to the side chain of asparagine 20. From the crystal structure of 5K, the serine 24 is expected to form a hydrogen bond to the carboxyl of asparagine 20. In addition to the differences between the hydrogen bonds expected from the crystal structure, there are some additional differences between the hydrogen bonds observed by NMR for RNase Sa and 5K. The amide of tyrosine 55 is observed to form a hydrogen bond to the carbonyl of isoleucine 70 in RNase Sa, but is not seen in 5K. Similarly, a

hydrogen bond from serine 90 to tyrosine 81 is observed in RNase Sa but not in 5K. Since these two hydrogen bonds are strand-to-strand interactions in the core of the protein and it is likely that they are formed in 5K but not observed in the NMR experiment, either from poor signal due to the solubility of 5K or because the hydrogen bond distance is slightly longer and has a weaker J-coupling.

Energetic Analysis

Table 27 shows the volumes calculated for residues in RNase Sa and the 5K variant. As discussed above, since water was excluded from the calculation, the Voronoi calculation assigns an infinite volume for some surface atoms. Residues that contain these surface atoms with infinite volumes were excluded from the table. In general, the residues in the 5K variant have larger volumes that the corresponding atom in RNase Sa, leading to the conclusion that RNase Sa has a higher packing density than the 5K variant and that van der Waals interactions would be stronger.

Table 27: Residues volumes in RNase Sa and the 5K variant

		RNase Sa			5K Variant			
		BB	SC	Total	BB	SC	Total	ΔV
LEU	8	57	166	224	59	172	231	7
LEU	11	80	129	209	80	142	222	13
ALA	**15**	**51**	**39**	90	**52**	**39**	**92**	**1**
THR	18	53	78	131	53	79	132	1
LEU	19	60	127	187	59	130	189	2
ILE	22				79	172	251	
GLY	26	95		95	86		86	-9
PHE	37	61	167	229	62	173	235	6
ASN	39	68	92	160	70	92	162	2
VAL	43				65	129	194	
LEU	44	60	104	164	54	110	165	1
TYR	52	55	153	208	55	153	208	0
GLU	54	51	139	190	53	151	204	15
TYR	55	53	192	245	56	202	258	13
THR	56	62	66	128	66	72	137	9
VAL	**57**	**54**	**83**	137	**54**	**83**	**136**	**-1**
ARG	65	76	218	294				
GLY	66	114		114	107		107	-7
ARG	69	50	148	198	53	166	219	20
ILE	**70**	**50**	**108**	157	**52**	**109**	**161**	**4**
ILE	**71**	**48**	**102**	150	**50**	**102**	**153**	**3**
THR	72	52	103	156	54	101	154	-1
GLY	73	88		88	103		103	15
GLU	78	60	140	200	62	131	193	-7
ASP	79	57	105	162	55	108	163	1
TYR	80	52	162	214	53	198	251	38
THR	82	52	60	112	52	63	115	4
TYR	86	67	195	262				
PHE	89	60	140	201	59	141	200	-1
ILE	92	79	109	188	70	114	184	-4
Total								123

Volumes are listed in $Å^3$. Residues in bold are completely buried.

$\Delta V = Volume_{5K} - Volume_{RNase\ Sa}$

Table 28: Energetic analysis of the native state of RNase Sa and the 5K variant

	RNase Sa	5K Variant	Difference	ΔΔG
Electrostatic Interactions[1]	9.3 kcal/mol	3.2 kcal/mol	-6.1 kcal/mol	6.1 kcal/mol
Hydrogen Bonding[2]	73 H bonds	74 H bonds	1 H bond	1 kcal/mol
Hydrophobic Burial[3]	276 -CH_2- groups	284 -CH_2- groups	8 -CH_2- groups	8 kcal/mol
Total				15.1 kcal/mol

[1] Calculated using Coulomb's Law as described previously (Huyghues-Despointes et al. 2003) with dielectric of 20
[2] Contribution to ΔΔG was estimated using 1 kcal/mol per hydrogen bond
[3] Calculated using pfis as in (Hebert et al. 1998), contribution to ΔΔG was estimated using 1 kcal/mol per -CH_2- group buried

Table 28 shows an estimate of the ΔΔG for the 5K variant at pH 7 based on electrostatic interactions, hydrogen bonding and the burial of hydrophobic surface area. Based on these values the 5K variant should have a ΔΔG of 15 kcal/mol, but the measured ΔΔG is –0.6 kcal/mol (Shaw et al. 2001). An understanding of the contributions to the free energy by the differences in packing, polar group burial and the denatured state are needed to fully account for the measured ΔΔG of the 5K variant.

Conclusion

The structures of RNase Sa and the 5K variant are very similar. The D1K mutation in 5K results in a slight rearrangement of the N-terminal residues to improve the strand-to-strand interactions between the first and last beta strands in RNase Sa. The E41K and E74K mutations appear to result in a minor rearrangement of the major loop. The E41K mutation results in a rearrangement of the side chain packing between the 57 to 78 loop

and the first turn (31 to 34) in the major loop. The E74K mutation appears to affect the packing at the end of the major helix (47 to 52) and results in a rearrangement of the backbone in that region. Overall the changes are slight and do not represent a major change in the structure.

SUMMARY

Correctly folded proteins are required for the proper function of all metabolic processes. Protein misfolding can lead to a variety of diseases including cystic fibrosis and over 20 different amyloid diseases. Understanding the forces involved in protein folding is essential to accurately predicting protein structure and understanding the various protein misfolding diseases.

When proteins fold, a significant amount of both polar and hydrophobic surface area is buried in the interior of the protein (Lesser and Rose 1990). It has long been known that the burial of hydrophobic surface area makes a significant and favorable contribution to protein stability (Kauzmann 1959), but the effect of polar group burial is not understood. While theoretical evidence suggests polar group burial is destabilizing (Honig and Yang 1995), experimental results show the removal of buried polar groups to be destabilizing (Pace 2001). One study replaced sixteen tyrosines in RNase Sa and Sa3 with phenylalanine and found the mutations to be generally destabilizing, even when the tyrosine side chain formed no hydrogen bonds (Pace et al. 2001). There was also a wide range of effects on protein stability; replacing tyrosine 52 in RNase Sa with phenylalanine reduced the stability by 3.6 kcal/mol while replacing tyrosine 33 in RNase Sa3 increased the stability by 0.5 kcal/mol. To understand the different effects on stability, the contribution of the tyrosine hydroxyl to the stability was determined by calculating the electrostatic and van der Waals interactions with the rest of the protein. I found the variation in the contribution of the tyrosine to the protein stability could be

explained by differences in the favorable van der Waals interactions of the tyrosine hydroxyl. These data also suggested that hydrogen bonding increased packing density as hydrogen bonded tyrosine had more atoms within 3.5 Å of the hydroxyl.

To further investigate the electrostatic and van der Waals interactions in proteins, I calculated all the non-bonded interactions in over 900 proteins. Atom groups were classified into five categories (aliphatic, aromatic, polar carbon, polar and alpha carbon). The data were analyzed to look at interactions between the five categories, interactions between each residue and the five categories and between each atom group and the five categories. Some interesting interactions were found, such as aromatic-aromatic stacking interactions appear to be dominated by interactions involving tryptophan. In the future, it would be useful to investigate residue-to-residue interactions in a similar fashion.

When calculating the non-bonded interactions for the tyrosine, I observed that hydrogen-bonded tyrosine had more atoms within 3.5 Å. Others have suggested that polar groups in the interior of the protein have a higher density (Kuntz 1972) and this has been suggested as a possible mechanism in which hydrogen bonding could increase the stability of folded proteins (Honig 1999). To further investigate the possibility that hydrogen bonds increase packing density, I determined the residue volumes and local packing density for nearly 900 high-resolution X-ray structures. Analyzing for the effect of hydrogen bonding and secondary structure showed that hydrogen bonds increased the local packing density and reduced residue volume by approximately 1.5 $Å^3$ per hydrogen

bond. Future work will include investigations of hydrophobic regions to identify how the local environment of hydrophobic groups affects packing density.

The final study is a structural comparison between RNase Sa and a previously reported variant (5K) in which five acidic groups were replaced with lysine (Pace et al. 2000). The 5K variant effectively reverses the net charge on the protein, but the positions were chosen to be surface exposed and non-interacting. Through a detailed structural analysis of crystallographic data combined with analysis of hydrogen bonding patterns using trans hydrogen bond J-coupling techniques, I show there are only minor perturbations in the structure of the 5K variant. These small perturbations are localized to some loop regions of the structure and have very little impact on the protein core.

REFERENCES

Acampora, G., and Hermans, J., Jr. 1967. Reversible denaturation of sperm whale myoglobin. I. Dependence on temperature, pH, and composition. *J Am Chem Soc* **89:** 1543-1547.

Albeck, S., Unger, R., and Schreiber, G. 2000. Evaluation of direct and cooperative contributions towards the strength of buried hydrogen bonds and salt bridges. *J Mol Biol* **298:** 503-520.

Anderson, D.E., Becktel, W.J., and Dahlquist, F.W. 1990. pH-induced denaturation of proteins: a single salt bridge contributes 3-5 kcal/mol to the free energy of folding of T4 lysozyme. *Biochemistry* **29:** 2403-2408.

Anfinsen, C.B. 1973. Principles that govern the folding of protein chains. *Science* **181:** 223-230.

Anson, M. 1945. Protein denaturation and the properties of protein groups. *Adv Protein Chem* **2:** 361-386.

Anson, M., and Mirsky, A.E. 1931. The reversibility of protein coagulation. *J Phys Chem* **35:** 185-193.

Bacova, M., Zelinkova, E., and Zelinka, J. 1971. Exocellular ribonuclease from *Streptomyces aureofaciens*. I. Isolation and purification. *Biochim Biophys Acta* **235:** 335-342.

Baldwin, R.L. 1986. Temperature dependence of the hydrophobic interaction in protein folding. *Proc Natl Acad Sci U S A* **83:** 8069-8072.

Barlow, D.J., and Thornton, J.M. 1983. Ion-pairs in proteins. *J Mol Biol* **168:** 867-885.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. 2000. The Protein Data Bank. *Nucleic Acids Res* **28:** 235-242.

Bernal, J.D. 1939. Structure of proteins. *Nature* **143:** 663-667.

Bhat, M.G., Ganley, L.M., Ledman, D.W., Goodman, M.A., and Fox, R.O. 1997. Stability studies of amino acid substitutions at tyrosine 27 of the staphylococcal nuclease beta-barrel. *Biochemistry* **36:** 12167-12174.

Blake, P.R., Park, J.B., Adams, M.W.W., and Summers, M.F. 1992. Novel observation of Nh...S(Cys) hydrogen-bond-mediated scalar coupling in Cd-113-substituted rubredoxin from *Pyrococcus furiosus*. *J Am Chem Soc* **114:** 4931-4933.

Bolen, D.W., and Santoro, M.M. 1988. Unfolding free energy changes determined by the linear extrapolation method. 2. Incorporation of delta G degrees N-U values in a thermodynamic cycle. *Biochemistry* **27:** 8069-8074.

Brady, J.E., and Humiston, G.E. 1986. *General chemistry principles and structure*, 4 ed. John Wiley & Sons, New York.

Brandts, J. 1964. The thermodynamics of protein denaturation. II. A model of reversible denaturation and interpretations regarding the stability of chymotrypsinogen. *J Am Chem Soc* **86:** 4302-4314.

Brandts, J., and Lumry, R. 1963. The reversible thermal denaturation of chymotrypsinogen. I. Experimental characterization. *J Phys Chem* **67:** 1484-1494.

Brandts, J.F., Oliveira, R.J., and Westort, C. 1970. Thermodynamics of protein denaturation. Effect of pressure on the denaturation of ribonuclease A. *Biochemistry* **9:** 1038-1047.

Byrne, M.P., Manuel, R.L., Lowe, L.G., and Stites, W.E. 1995. Energetic contribution of side chain hydrogen bonding to the stability of staphylococcal nuclease. *Biochemistry* **34:** 13949-13960.

Chen, Y.W., Fersht, A.R., and Henrick, K. 1993. Contribution of buried hydrogen bonds to protein stability. The crystal structures of two barnase mutants. *J Mol Biol* **234:** 1158-1170.

Chothia, C. 1975. Structural invariants in protein folding. *Nature* **254:** 304-308.

Cieplak, P., Caldwell, J., and Kollman, P. 2001. Molecular mechanical models for organic and biological systems going beyond the atom centered two body additive approximation: aqueous solution free energies of methanol and N-methyl acetamide, nucleic acid base, and amide hydrogen bonding and chloroform/water partition coefficients of the nucleic acid bases. *J Comp Chem* **22:** 1048-1057.

Cordier, F., and Grzesiek, S. 1999. Direct observation of hydrogen bonds in proteins by interresidue $^{3h}J_{NC'}$ scalar couplings. *J Am Chem Soc* **121:** 1601-1602.

Cordier, F., Wang, C.Y., Grzesiek, S., and Nicholson, L.K. 2000. Ligand-induced strain in hydrogen bonds of the c-Src SH3 domain detected by NMR. *J Mol Biol* **304:** 497-505.

Cornilescu, G., Hu, J.S., and Bax, A. 1999a. Identification of the hydrogen bonding network in a protein by scalar couplings. *J Am Chem Soc* **121:** 2949-2950.

Cornilescu, G., Ramirez, B.E., Frank, M.K., Clore, G.M., Gronenborn, A.M., and Bax, A. 1999b. Correlation between $^{3h}J_{NC'}$ and hydrogen bond length in proteins. *J Am Chem Soc* **121:** 6275-6279.

Creamer, T.P., and Rose, G.D. 1992. Side-chain entropy opposes alpha-helix formation but rationalizes experimentally determined helix-forming propensities. *Proc Natl Acad Sci U S A* **89:** 5937-5941.

Creighton, T.E. 1997. *Proteins: structures and molecular properties*, 2nd ed. W. H. Freeman and Company, New York.

D'Aquino, J.A., Gomez, J., Hilser, V.J., Lee, K.H., Amzel, L.M., and Freire, E. 1996. The magnitude of the backbone conformational entropy change in protein folding. *Proteins* **25:** 143-156.

Dang, L.X., Pearlman, D.A., and Kollman, P.A. 1990. Why do A.T base pairs inhibit Z-DNA formation? *Proc Natl Acad Sci U S A* **87:** 4630-4634.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. 1995. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6:** 277-293.

Dill, K.A. 1990. Dominant forces in protein folding. *Biochemistry* **29:** 7133-7155.

Dingley, A.J., and Grzesiek, S. 1998. Direct observation of hydrogen bonds in nucleic acid base pairs by internucleotide (2)J(NN) couplings. *J Am Chem Soc* **120:** 8293-8297.

Dixon, R.W., and Kollman, P.A. 1997. Advancing beyond the atom-centered model in additive and nonadditive molecular mechanics. *J Comp Chem* **18:** 1632-1646.

Doig, A.J., and Sternberg, M.J. 1995. Side-chain conformational entropy in protein folding. *Protein Sci* **4:** 2247-2251.

Dougherty, D.A. 1996. Cation-pi interactions in chemistry and biology: a new view of benzene, Phe, Tyr, and Trp. *Science* **271:** 163-168.

Dunfield, L.G., Burgess, A.W., and Scheraga, H.A. 1978. Energy parameters in polypeptides. 8. Empirical potential-energy algorithm for conformational-analysis of large molecules. *J Phys Chem* **82:** 2609-2616.

Edsall, J.T. 1935. Apparent molal heat capacities of amino acids and other organic compounds. *J Am Chem Soc* **57:** 1506-1507.

Eisenberg, D., and Kauzmann, W. 1969. *The structure and properties of water*. Oxford University Press, Oxford.

Elcock, A.H. 1999. Realistic modeling of the denatured states of proteins allows accurate calculations of the pH dependence of protein stability. *J Mol Biol* **294:** 1051-1062.

Fersht, A.R., Shi, J.P., Knill-Jones, J., Lowe, D.M., Wilkinson, A.J., Blow, D.M., Brick, P., Carter, P., Waye, M.M., and Winter, G. 1985. Hydrogen bonding and biological specificity analysed by protein engineering. *Nature* **314:** 235-238.

Finney, J.L. 1970. Random packings and structure of simple liquids. 1. Geometry of random close packing. *Proc R Soc Lon Ser-A* **319:** 479-493.

Fleischman, S.H., and Brooks, C.L., III. 1987. Thermodynamics of aqueous solvation - solution properties of alcohols and alkanes. *J Chem Phys* **87:** 3029-3037.

Flory, P.J. 1956. Theory of elastic mechanisms in fibrous proteins. *J Am Chem Soc* **78:** 5222-5235.

Gallivan, J.P., and Dougherty, D.A. 1999. Cation-pi interactions in structural biology. *Proc Natl Acad Sci U S A* **96:** 9459-9464.

Garrett, D., Powers, R., Gronenborn, A.M., and Clore, G.M. 1991. A common sense approach to peak picking in two-, three-, and four-dimensional spectra using automatic computer analysis of contour diagrams. *J Magn Reson* **95:** 214-220.

Gerstein, M. 1992. A resolution-sensitive procedure for comparing protein surfaces and its application to the comparison of antigen-combining sites. *Acta Crystallogr A* **48:** 271-276.

Gerstein, M., Tsai, J., and Levitt, M. 1995. The volume of atoms on the protein surface: calculated from simulation, using Voronoi polyhedra. *J Mol Biol* **249:** 955-966.

Giletto, A., and Pace, C.N. 1999. Buried, charged, non-ion-paired aspartic acid 76 contributes favorably to the conformational stability of ribonuclease T-1. *Biochemistry* **38:** 13379-13384.

Gilli, G., and Gilli, P. 2000. Towards an unified hydrogen-bond theory. *J Mol Struct* **552:** 1-15.

Gilson, M.K., and Honig, B.H. 1988. Energetics of charge-charge interactions in proteins. *Proteins* **3:** 32-52.

Haber, E., and Anfinsen, C.B. 1962. Side-chain interactions governing the pairing of half-cystine residues in Ribonuclease. *J Biol Chem* **237:** 1839-1844.

Hagler, A.T., Lifson, S., and Dauber, P. 1979. Consistent force-field studies of intermolecular forces in hydrogen-bonded crystals. 2. Benchmark for the objective comparison of alternative force-fields. *J Am Chem Soc* **101:** 5122-5130.

Hamill, S.J., Cota, E., Chothia, C., and Clarke, J. 2000. Conservation of folding and stability within a protein family: the tyrosine corner as an evolutionary cul-de-sac. *J Mol Biol* **295:** 641-649.

Harpaz, Y., Gerstein, M., and Chothia, C. 1994. Volume changes on protein-folding. *Structure* **2:** 641-649.

Harrington, W.F., and Sela, M. 1959. A comparison of the physical chemical properties of oxidized and reduced alkylated ribonuclease. *Biochim Biophys Acta* **31:** 427-434.

Harris, T.K., Zhao, Q., and Mildvan, A.S. 2000. NMR studies of strong hydrogen bonds in enzymes and in a model compound. *J Mol Struct* **552:** 97-109.

Hebert, E.J., Giletto, A., Sevcik, J., Urbanikova, L., Wilson, K.S., Dauter, Z., and Pace, C.N. 1998. Contribution of a conserved asparagine to the conformational stability of ribonucleases Sa, Ba, and T1. *Biochemistry* **37:** 16192-16200.

Hebert, E.J., Grimsley, G.R., Hartley, R.W., Horn, G., Schell, D., Garcia, S., Both, V., Sevcik, J., and Pace, C.N. 1997. Purification of ribonucleases Sa, Sa2, and Sa3 after expression in *Escherichia coli*. *Protein Expr Purif* **11:** 162-168.

Hendsch, Z.S., and Tidor, B. 1994. Do salt bridges stabilize proteins? A continuum electrostatic analysis. *Protein Sci* **3:** 211-226.

Honig, B. 1994. Free energy balance in protein folding. *J Mol Biol* **237:** 602-614.

Honig, B. 1999. Protein folding: from the levinthal paradox to structure prediction. *J Mol Biol* **293:** 283-293.

Honig, B., and Yang, A.S. 1995. Free energy balance in protein folding. *Adv Protein Chem* **46:** 27-58.

Hubbard, S.J., Gross, K.H., and Argos, P. 1994. Intramolecular cavities in globular proteins. *Protein Eng* **7:** 613-626.

Huyghues-Despointes, B.M., Klingler, T.M., and Baldwin, R.L. 1995. Measuring the strength of side-chain hydrogen bonds in peptide helices: the Gln.Asp (i, i + 4) interaction. *Biochemistry* **34:** 13267-13271.

Huyghues-Despointes, B.M., Thurlkill, R.L., Daily, M.D., Schell, D., Briggs, J.M., Antosiewicz, J.M., Pace, C.N., and Scholtz, J.M. 2003. pK values of histidine residues in ribonuclease Sa: effect of salt and net charge. *J Mol Biol* **325:** 1093-1105.

Jorgensen, W.L., Maxwell, D.S., and Tiradorives, J. 1996. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J Am Chem Soc* **118:** 11225-11236.

Kabsch, W., and Sander, C. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22:** 2577-2637.

Kauzmann, W. 1954. Denaturation of proteins and enzymes. In *The machanism of enzyme action*. (eds. W.D. McElroy, and B. Glass), pp. 70-120. Johns Hopkins University Press, Baltimore.

Kauzmann, W. 1959. Some Factors in the interpretation of protein denaturation. *Adv Protein Chem* **14:** 1-63.

Kazmirski, S.L., Wong, K.B., Freund, S.M., Tan, Y.J., Fersht, A.R., and Daggett, V. 2001. Protein folding from a highly disordered denatured state: the folding pathway of chymotrypsin inhibitor 2 at atomic resolution. *Proc Natl Acad Sci U S A* **98:** 4349-4354.

Koh, J.T., Cornish, V.W., and Schultz, P.G. 1997. An experimental approach to evaluating the role of backbone interactions in proteins using unnatural amino acid mutagenesis. *Biochemistry* **36:** 11314-11322.

Kono, H., Saito, M., and Sarai, A. 2000. Stability analysis for the cavity-filling mutations of the Myb DNA-binding domain utilizing free-energy calculations. *Proteins* **38:** 197-209.

Kraulis, P.J. 1991. MOLSCRIPT: a program to product both detailed and schematic plots of protein structures. *Journal of Applied Crystallography* **24:** 946-950.

Kuntz, I.D. 1972. Tertiary structure in carboxypeptidase. *J Am Chem Soc* **94:** 8568-8572.

Latimer, W.M., and Rodebush, W.H. 1920. Polarity and ionization from the standpoint of the Lewis theory of valence. *J Am Chem Soc* **42:** 1419-1433.

Laurents, D., Perez-Canadillas, J.M., Santoro, J., Rico, M., Schell, D., Pace, C.N., and Bruix, M. 2001. Solution structure and dynamics of ribonuclease Sa. *Proteins* **44:** 200-211.

Laurents, D.V., Huyghues-Despointes, B.M., Bruix, M., Thurlkill, R.L., Schell, D., Newsom, S., Grimsley, G.R., Shaw, K.L., Trevino, S., Rico, M., et al. 2003. Charge-charge interactions are key determinants of the pK values of ionizable groups in ribonuclease Sa (pI=3.5) and a basic variant (pI=10.2). *J Mol Biol* **325:** 1077-1092.

Laurents, D.V., Perez-Canadillas, J.M., Santoro, J., Rico, M., Schell, D., Hebert, E.J., Pace, C.N., and Bruix, M. 1999. Letter to the editor: sequential assignment and solution secondary structure of doubly labelled ribonuclease Sa. *J Biomol NMR* **14:** 89-90.

Lazaridis, T., Archontis, G., and Karplus, M. 1995. Enthalpic contribution to protein stability: insights from atom-based calculations and statistical mechanics. *Adv Protein Chem* **47:** 231-306.

Lee, K.H., Xie, D., Freire, E., and Amzel, L.M. 1994. Estimation of changes in side chain configurational entropy in binding and folding: general methods and application to helix formation. *Proteins* **20:** 68-84.

Lennard-Jones, J.E. 1931. Cohesion. *The Proceedings of the Physical Society* **43:** 461-482.

Lesser, G.J., and Rose, G.D. 1990. Hydrophobicity of amino acid subgroups in proteins. *Proteins* **8:** 6-13.

Lifson, S., Hagler, A.T., and Dauber, P. 1979. Consistent force-field studies of intermolecular forces in hydrogen-bonded crystals. 1. Carboxylic-acids, amides, and the C=O...H- hydrogen-bonds. *J Am Chem Soc* **101:** 5111-5121.

Linderstrom-Lang, K. 1924. On the ionisation of proteins. *C R Trav Lab Carlsberg* **15:** 1-29.

Liu, A., Hu, W., Qamar, S., and Majumdar, A. 2000. Sensitivity enhanced NMR spectroscopy by quenching scalar coupling mediated relaxation: application to the direct observation of hydrogen bonds in 13C/15N-labeled proteins. *J Biomol NMR* **17:** 55-61.

Lomize, A.L., Reibarkh, M.Y., and Pogozheva, I.D. 2002. Interatomic potentials and solvation parameters from protein engineering data for buried residues. *Protein Sci* **11:** 1984-2000.

London, F. 1937. The general theory of molecular forces. *Transactions of the Faraday Society* **33:** 8-26.

Lumry, R., and Eyring, H. 1954. Conformation changes of proteins. *J Phys Chem* **58:** 110-120.

Mackerell, A.D., Bashford, D., Bellott, M., Dunbrack, R.L., Evanseck, J.D., Field, M.J., Fischer, S., Gao, J., Guo, H., Ha, S., et al. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* **102:** 3586-3616.

Makhatadze, G.I., and Privalov, P.L. 1995. Energetics of protein structure. *Adv Protein Chem* **47:** 307-425.

Martensson, L.G., Jonsson, B.H., Andersson, M., Kihlgren, A., Bergenhem, N., and Carlsson, U. 1992. Role of an evolutionarily invariant serine for the stability of human carbonic anhydrase II. *Biochim Biophys Acta* **1118:** 179-186.

Matthews, B.W. 1995. Studies on protein stability with T4 lysozyme. *Adv Protein Chem* **46:** 249-278.

Matthews, B.W., Nicholson, H., and Becktel, W.J. 1987. Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding. *Proc Natl Acad Sci U S A* **84:** 6663-6667.

McDonald, I.K., and Thornton, J.M. 1994. Satisfying hydrogen bonding potential in proteins. *J Mol Biol* **238:** 777-793.

Meng, E.C., Cieplak, P., Caldwell, J.W., and Kollman, P.A. 1994. Accurate solvation free-energies of acetate and methylammonium ions calculated with a polarizable water model. *J Am Chem Soc* **116:** 12061-12062.

Mirsky, A.E., and Pauling, L. 1936. On the structure of native, denatured, and coagulated proteins. *Proc Natl Acad Sci U S A* **22:** 439-447.

Mitchell, J.B.O.P., S. L. 1990. The nature of the N-H···O=C hydrogen bond: an intermolecular perturbation theory study of the formamide/formaldehyde complex. *J Comp Chem* **11:** 1217-1233.

Momany, F.A., Carruthe.Lm, Mcguire, R.F., and Scheraga, H.A. 1974. Intermolecular potentials from crystal data. 3. Determination of empirical potentials and

application to packing configurations and lattice energies in crystals of hydrocarbons, carboxylic-acids, amines, and amides. *J Phys Chem* **78:** 1595-1620.

Myers, J.K., and Pace, C.N. 1996. Hydrogen bonding stabilizes globular proteins. *Biophys J* **71:** 2033-2039.

Myers, J.K., Pace, C.N., and Scholtz, J.M. 1995. Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci* **4:** 2138-2148.

Pace, C.N. 1975. The stability of globular proteins. *CRC Crit Rev Biochem* **3:** 1-43.

Pace, C.N. 2001. Polar group burial contributes more to protein stability than nonpolar group burial. *Biochemistry* **40:** 310-313.

Pace, C.N., Alston, R.W., and Shaw, K.L. 2000. Charge-charge interactions influence the denatured state ensemble and contribute to protein stability. *Protein Sci* **9:** 1395-1398.

Pace, C.N., Grimsley, G.R., Thomson, J.A., and Barnett, B.J. 1988. Conformational stability and activity of ribonuclease T1 with zero, one, and two intact disulfide bonds. *J Biol Chem* **263:** 11820-11825.

Pace, C.N., Hebert, E.J., Shaw, K.L., Schell, D., Both, V., Krajcikova, D., Sevcik, J., Wilson, K.S., Dauter, Z., Hartley, R.W., et al. 1998. Conformational stability and thermodynamics of folding of ribonucleases Sa, Sa2 and Sa3. *J Mol Biol* **279:** 271-286.

Pace, C.N., Horn, G., Hebert, E.J., Bechert, J., Shaw, K., Urbanikova, L., Scholtz, J.M., and Sevcik, J. 2001. Tyrosine hydrogen bonds make a large contribution to protein stability. *J Mol Biol* **312:** 393-404.

Pace, C.N., Shirley, B.A., McNutt, M., and Gajiwala, K. 1996. Forces contributing to the conformational stability of proteins. *Faseb J* **10:** 75-83.

Pauling, L. 1928. The shared-electron chemical bond. *Proc Natl Acad Sci U S A* **14:** 359-362.

Pauling, L., and Corey, R.B. 1951. The pleated sheet, a new layer configuration of polypeptide chains. *Proc Natl Acad Sci U S A* **37:** 251-256.

Pauling, L., Corey, R.B., and Branson, H.R. 1951. The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain. *Proc Natl Acad Sci U S A* **37:** 205-211.

Perutz, M.F., and Raidt, H. 1975. Stereochemical basis of heat stability in bacterial ferredoxins and in haemoglobin A2. *Nature* **255:** 256-259.

Ponder, J.W. 2001. TINKER: software tools for molecular design, 3.9 ed. Washington University School of Medicine, Saint Louis.

Pontius, J., Richelle, J., and Wodak, S.J. 1996. Deviations from standard atomic volumes as a quality measure for protein crystal structures. *J Mol Biol* **264:** 121-136.

Prevost, M., Wodak, S.J., Tidor, B., and Karplus, M. 1991. Contribution of the hydrophobic effect to protein stability: analysis based on simulations of the Ile-96----Ala mutation in barnase. *Proc Natl Acad Sci U S A* **88:** 10880-10884.

Privalov, P.L. 1979. Stability of proteins: small globular proteins. *Adv Protein Chem* **33:** 167-241.

Privalov, P.L. 1989. Thermodynamic problems of protein-structure. *Annu Rev Biophys Bio* **18:** 47-69.

Richards, F.M. 1974. The interpretation of protein structures: total volume, group volume distributions and packing density. *J Mol Biol* **82:** 1-14.

Sali, D., Bycroft, M., and Fersht, A.R. 1991. Surface electrostatic interactions contribute little of stability of barnase. *J Mol Biol* **220:** 779-788.

Santoro, M.M., and Bolen, D.W. 1988. Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. *Biochemistry* **27:** 8063-8068.

Serrano, L., Kellis, J.T., Jr., Cann, P., Matouschek, A., and Fersht, A.R. 1992. The folding of an enzyme. II. Substructure of barnase and the contribution of different interactions to protein stability. *J Mol Biol* **224:** 783-804.

Sevcik, J., Dauter, Z., Lamzin, V.S., and Wilson, K.S. 1996. Ribonuclease from *Streptomyces aureofaciens* at atomic resolution. *Acta Crystallogr D Biol Crystallogr* **52:** 327-344.

Sevcik, J., Hill, C.P., Dauter, Z., and Wilson, K.S. 1993. Complex of ribonuclease from *Streptomyces aureofaciens* with 2'-Gmp at 1.7-angstrom resolution. *Acta Crystallogr D Biol Crystallogr* **49:** 257-271.

Sevcik, J., Lamzin, V.S., Dauter, Z., and Wilson, K.S. 2002a. Atomic resolution data reveal flexibility in the structure of RNase Sa. *Acta Crystallogr D Biol Crystallogr* **58:** 1307-1313.

Sevcik, J., Urbanikova, L., Leland, P.A., and Raines, R.T. 2002b. X-ray structure of two crystalline forms of a streptomycete ribonuclease with cytotoxic activity. *J Biol Chem* **277:** 47325-47330.

Shaw, K.L., Grimsley, G.R., Yakovlev, G.I., Makarov, A.A., and Pace, C.N. 2001. The effect of net charge on the solubility, activity, and stability of ribonuclease Sa. *Protein Sci* **10:** 1206-1215.

Shire, S.J., Hanania, G.I.H., and Gurd, F.R.N. 1974. Electrostatic effects in myoglobin - hydrogen-ion equilibria in sperm whale ferrimyoglobin. *Biochemistry* **13:** 2967-2973.

Shirley, B.A., Stanssens, P., Hahn, U., and Pace, C.N. 1992. Contribution of hydrogen bonding to the conformational stability of ribonuclease T1. *Biochemistry* **31:** 725-732.

Shortle, D. 1996. The denatured state (the other half of the folding equation) and its role in protein stability. *Faseb J* **10:** 27-34.

Sippl, M.J., Ortner, M., Jaritz, M., Lackner, P., and Flockner, H. 1996. Helmholtz free energies of atom pair interactions in proteins. *Fold Des* **1:** 289-298.

Sneddon, S.F., and Tobias, D.J. 1992. The role of packing interactions in stabilizing folded proteins. *Biochemistry* **31:** 2842-2846.

Spolar, R.S., and Record, M.T., Jr. 1994. Coupling of local folding to site-specific binding of proteins to DNA. *Science* **263:** 777-784.

Stickle, D.F., Presta, L.G., Dill, K.A., and Rose, G.D. 1992. Hydrogen bonding in globular proteins. *J Mol Biol* **226:** 1143-1159.

Stillinger, F.H. 1980. Water revisited. *Science* **209:** 451-457.

Sturtevant, J.M. 1977. Heat-capacity and entropy changes in processes involving proteins. *Proc Natl Acad Sci U S A* **74:** 2236-2240.

Sugita, Y., and Kitao, A. 1998. Dependence of protein stability on the structure of the denatured state: free energy calculations of I56V mutation in human lysozyme. *Biophys J* **75:** 2178-2187.

Sun, D.P., Sauer, U., Nicholson, H., and Matthews, B.W. 1991. Contributions of engineered surface salt bridges to the stability of T4 lysozyme determined by directed mutagenesis. *Biochemistry* **30:** 7142-7153.

Takano, K., Yamagata, Y., Kubota, M., Funahashi, J., Fujii, S., and Yutani, K. 1999. Contribution of hydrogen bonds to the conformational stability of human lysozyme: calorimetry and X-ray analysis of six Ser --> Ala mutants. *Biochemistry* **38:** 6623-6629.

Tanford, C. 1962. Contribution of hydrophobic interactions to the stability of the globular conformation of proteins. *J Am Chem Soc* **84:** 4240-4247.

Tanford, C., and Kirkwood, J.G. 1957. Theory of protein titration curves. I. General equations for impenetrable spheres. *J Am Chem Soc* **79:** 5333-5339.

Tanner, J.J., Hecht, R.M., and Krause, K.L. 1996. Determinants of enzyme thermostability observed in the molecular structure of *Thermus aquaticus* D-glyceraldehyde-3-phosphate dehydrogenase at 2.5 angstroms resolution. *Biochemistry* **35:** 2597-2609.

Tidor, B. 1990. Simulation analysis of the stability mutant R96H of T4 lysozyme. *Proc Natl Acad Sci U S A* **87:** 8481-8485.

Tsai, C.J., Maizel, J.V., Jr., and Nussinov, R. 2002. The hydrophobic effect: a new insight from cold denaturation and a two-state water structure. *Crit Rev Biochem Mol Biol* **37:** 55-69.

Tsai, J., and Gerstein, M. 2002. Calculations of protein volumes: sensitivity analysis and parameter database. *Bioinformatics* **18:** 985-995.

Tsai, J., Taylor, R., Chothia, C., and Gerstein, M. 1999. The packing density in proteins: standard radii and volumes. *J Mol Biol* **290:** 253-266.

Voronoi, G.F. 1908. Nouveles applications des paramétres continus à la théorie des formes quadratiques. *J. Reine Angew. Math.* **134:** 198-287.

Wang, G., and Dunbrack, R.L. 2002. PISCES: a protein sequence culling server. *Bioinformatics*.

Wang, J.M., Cieplak, P., and Kollman, P.A. 2000. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J Comp Chem* **21:** 1049-1074.

Yamagata, Y., Kubota, M., Sumikawa, Y., Funahashi, J., Takano, K., Fujii, S., and Yutani, K. 1998. Contribution of hydrogen bonds to the conformational stability of human lysozyme: calorimetry and X-ray analysis of six tyrosine --> phenylalanine mutants. *Biochemistry* **37:** 9355-9362.

Yang, A.S., and Honig, B. 1993. On the pH dependence of protein stability. *J Mol Biol* **231:** 459-474.

Zipp, A., and Kauzmann, W. 1973. Pressure denaturation of metmyoglobin. *Biochemistry* **12:** 4217-4228.

VITA

David Andrew Schell

1265 Colwell Ct

Bryan, Texas 77807

David Andrew Schell was born April 7, 1975 in Bryan, Texas. He lived in Plano, Texas from 1977 to 1982, when he moved to Alice Springs, Australia. In 1986, he returned from Australia and lived in Dallas, Texas until he graduated from Jesuit College Preparatory School in 1993. From there he went to Texas A&M University receiving a Bachelor of Science in computer science in May of 1997 and a Bachelor of Science in biochemistry and genetics in December of that same year. He entered graduate school at Texas A&M University in August, 1999 and received his Doctor of Philosophy in biochemistry in December, 2003.